

**TOSHKENT AXBOROT TEXNOLOGIYALARI UNIVERSITETI
HUZURIDAGI ILMIY DARAJALAR BERUVCHI
DSc.13/30.12.2019.T.07.01 RAQAMLI ILMIY KENGASH**

**ALISHER NAVOIY NOMIDAGI TOSHKENT DAVLAT
O‘ZBEK TILI VA ADABIYOTI UNIVERSITETI**

AXMEDOVA XOLISXON ILXOMOVNA

**O‘ZBEK TILIDAGI GAPLARNI SEMANTIK ANALIZ QILISHNING
MODELI, ALGORITMLARI VA AXBOROT TIZIMINI ISHLAB CHIQISH**

05.01.10–Axborot olish tizimlari va jarayonlari

**TEXNIKA FANLARI BO‘YICHA FALSAFA DOKTORI (PhD)
DISSERTATSIYASI AVTOREFERATI**

Toshkent – 2023

**Texnika fanlari bo'yicha falsafa doktori (PhD)
Dissertatsiyasi avtoreferati mundarijasi**

**Оглавление автореферата диссертации доктора философии (PhD)
по техническим наукам**

**Contents of dissertation abstract of doctor of philosophy (PhD)
on technical sciences**

Axmedova Xolisxon Ilxomovna

O'zbek tilidagi gaplarni semantik analiz qilishning
modeli, algoritmlari va axborot tizimini ishlab chiqish 3

Ахмедова Холисхон Илхомовна

Разработка модели, алгоритмов и информационной системы
семантического анализа предложений узбекского языка 22

Axmedova Xolisxon Ilxomovna

Development of a model, algorithms and information system
for semantic analysis of sentences in the Uzbek language 43

E'lon qilingan ishlar ro'yxati

Список опубликованных работ
List of published works 47

**TOSHKENT AXBOROT TEXNOLOGIYALARI UNIVERSITETI
HUZURIDAGI ILMIY DARAJALAR BERUVCHI
DSc.13/30.12.2019.T.07.01 RAQAMLI ILMIY KENGASH**

**ALISHER NAVOIY NOMIDAGI TOSHKENT DAVLAT
O‘ZBEK TILI VA ADABIYOTI UNIVERSITETI**

AXMEDOVA XOLISXON ILXOMOVNA

**O‘ZBEK TILIDAGI GAPLARNI SEMANTIK ANALIZ QILISHNING
MODELI, ALGORITMLARI VA AXBOROT TIZIMINI ISHLAB CHIQISH**

05.01.10–Axborot olish tizimlari va jarayonlari

**TEXNIKA FANLARI BO‘YICHA FALSAFA DOKTORI (PhD)
DISSERTATSIYASI AVTOREFERATI**

Texnika fanlari bo'yicha falsafa doktori (PhD) dissertatsiyasi mavzusi O'zbekiston Respublikasi Oliy ta'lim, fan va innovatsiyalar vazirligi huzuridagi Oliy attestatsiya komissiyasida B2023.1.PhD/T2289 raqam bilan ro'yxatga olingan.

Dissertatsiya Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universitetida bajarilgan.

Dissertatsiya avtoreferati uch tilda (o'zbek, rus, ingliz (rezyume)) Ilmiy kengash veb-sahifasida (www.tuit.uz) va «Ziyonet» axborot-ta'lim portalida (www.ziyonet.uz) joylashtirilgan.

Ilmiy rahbar:

Elov Botir Boltayevich

texnika fanlari bo'yicha falsafa doktori, dotsent

Rasmiy opponentlar:

Saidov Abdusobirjon Abduraxmonovich

texnika fanlari doktori, professor

Muxamedieva Dilnoz Tulqunovna

texnika fanlari doktori, professor

Yetakchi tashkilot:

Sharof Rashidov nomidagi Samarqand davlat universiteti

Dissertatsiya himoyasi Toshkent axborot texnologiyalari universiteti huzuridagi DSc.13/30.12.2019.T.07.01 raqamli Ilmiy kengashning 2023-yil «12» dekabr da soat 14:00 dagi majlisida bo'lib o'tadi. (Manzil: 100084, Toshkent shahri, Amir Temur ko'chasi, 108-uy. Tel.: (99871) 238-64-43; e-mail: iktuit@tuit.uz).

Dissertatsiya bilan Toshkent axborot texnologiyalari universitetining Axborot-resurs markazida tanishish mumkin (2857 raqam bilan ro'yxatga olingan). (Manzil: 100084, Toshkent shahri, Amir Temur ko'chasi, 108-uy. Tel.: (99871) 238-64-70).

Dissertatsiya avtoreferati 2023-yil «7» dekabr kuni tarqatildi.

(2023-yil «03» noyabr dagi 22 raqamli reyestr bayonnomasi).



M.M.Musayev

Ilmiy darajalar beruvchi ilmiy kengash raisi, texnika fanlari doktori, professor

N.O.Raximov

Ilmiy darajalar beruvchi ilmiy kengash ilmiy kotibi, texnika fanlari doktori, dotsent

Dj.B.Sultanov

Ilmiy darajalar beruvchi ilmiy kengash huzuridagi ilmiy seminar raisi, texnika fanlari doktori, dotsent

KIRISH (falsafa doktori (PhD) dissertasiyasi annotasiyasi)

Dissertasiya mavzusining dolzarbligi va zarurati. Jahonda kompyuter lingvistikasi sohasiga zamonaviy axborot texnologiyalarini keng joriy qilish orqali mavjud vazifalarni avtomatlashtirish, inson-mashina interfeysini soddalashtirish va qulaylashtirishga bo'lgan urinishlar natijasida XXI asrga kelib insonlar tabiiy tillarni o'rganish, ulardagi gaplarni analiz qilishni axborot tizimlari orqali amalga oshirilmoqda. Bunday vositalar so'nggi axborot-kommunikatsiya texnologiyalarining mahsuli bo'lib, tasvirlar va nutqqa texnik ishlov berish (processing) hamda tanib olish (recognition) tizimlari, matnlarni morfologik, sintaktik analiz qilish tizimlari shular jumlasidandir. Albatta kompyuter lingvistikasi sohasida zamonaviy texnologiyalar bo'yicha ilg'or bo'lgan AQSH, Fransiya, Germaniya, Buyuk Britaniya, Xitoy, Janubiy Koreya, Yaponiya, Rossiya kabi davlatlarning yetakchi olimlari olib borayotgan tadqiqotlari natijalariga katta e'tibor qaratilmoqda.

Jahonda turli tabiiy tillardagi gaplarni semantik analiz qilishni amalga oshiradigan ko'plab analizatorlar mavjud bo'lib, ularning ishlab chiqishda qo'llaniladigan usul va vositalar tanlangan tilning leksik va grammatik xususiyatlariga ko'ra turlichadir. Bu borada, matnlarga dastlabki ishlov berish va lingvistik analizatorlar ishlab chiqish bo'yicha ilmiy tadqiqotlar olib borish, tabiiy tilning o'ziga xos xususiyatlarini lingvistik jihatdan to'liq va mukammal ifodalashni amalga oshiruvchi axborot tizimlari ma'lumotlar bazasining ER (Entity Relationship) modelini ishlab chiqish, tilga ishlov berish biznes-jaryonlarini modellashtirish hamda axborot tizimlarini ishlab chiqishga erishish uchun maqbul yondashuvlar va yechimlarni tanlash muhim vazifalar hisoblanadi.

Respublikamizda zamonaviy axborot texnologiyalaridan keng foydalanish va ularni tadqiq etish, ishlab chiqilayotgan axborot tizimlarni ommalashtirish hamda soha bo'yicha olib borilayotgan tadqiqot ishlarini qo'llab-quvvatlash davlat darajasida amalga oshirilmoqda. Xususan, O'zbekiston Respublikasi Prezidentining "Raqamli O'zbekiston – 2030 strategiyasini tasdiqlash va uni samarali amalga oshirish chora-tadbirlari to'g'risida"gi Farmonida, jumladan, "...axborot texnologiyalari sohasidagi ustuvor yo'nalishlar bo'yicha fundamental va amaliy tadqiqotlarni olib borish ... istiqbolli innovatsion ishlanmalar va startap loyihalarni amalga oshirish hamda ularni tijoratlashtirish va texnologiyalar transferini qo'llab-quvvatlash ... raqamli texnologiyalar sohasida ilmiy-tadqiqot ishlarini rivojlantirish va rag'batlantirish, ularning tashkiliy mexanizmlarini takomillashtirish..."¹ kabi bir qator vazifalar belgilangan. Ushbu vazifalarni amalga oshirish, jumladan, bugungi kunning eng zamonaviy texnologiyalari asosida ishlab chiqarilayotgan murakkab tizimlarni tahlil qilish va ulardan foydalanishda inson-mashina interfeysining o'zbek modelini ishlab chiqish maqsadida turli loyihalar doirasida olib borilgan tadqiqotlar asosida axborot tizimlarni, xususan, o'zbek tilidagi gaplarni semantik analiz qilish axborot tizimlarini ishlab chiqish va foydalanishga taqdim etish bugungi kunning muhim vazifalardan biri hisoblanadi.

¹ O'zbekiston Respublikasi Prezidentining 2020 yil 5 oktyabrdagi PF-6079-son "Raqamli O'zbekiston – 2030 strategiyasini tasdiqlash va uni samarali amalga oshirish chora-tadbirlari to'g'risida"gi farmoni

Respublikamizda axborot texnologiyalari sohasini rivojlantirish hamda mavjud texnik va texnologik vositalarni takomillashtirish bo'yicha O'zbekiston Respublikasi Prezidentining 2017 yil 30-iyundagi PF-5099-son "Respublikada axborot texnologiyalari sohasini rivojlantirish uchun shart-sharoitlarni tubdan yaxshilash chora-tadbirlari to'g'risida"gi Farmoni, 2017 yil 20-apreldagi PQ-2909-son "Oliy ta'lim tizimini yanada rivojlantirish chora-tadbirlari to'g'risida"gi, 2020 yil 6-oktyabrdagi PQ-4851-son "Axborot texnologiyalari sohasida ta'lim tizimini yanada takomillashtirish, ilmiy tadqiqotlarni rivojlantirish va ularni IT-industriya bilan integratsiya qilish chora-tadbirlari to'g'risida"gi Qarorlari, 2020 yil 20-oktyabrdagi PF-6084-son "Mamlakatimizda o'zbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari to'g'risida" Farmoni hamda mazkur faoliyatga tegishli boshqa me'yoriy-huquqiy hujjatlarda belgilangan vazifalarni amalga oshirishda ushbu dissertatsiya tadqiqoti muayyan darajada xizmat qiladi.

Dissertasiyaning O'zbekiston Respublikasi fan va texnologiyalar taraqqiyoti ustuvor yo'nalishlariga mosligi. Mazkur tadqiqot ishi respublika fan va texnologiyalar rivojlanishining IV. "Axborotlashtirish va axborot-kommunikatsiya texnologiyalarini rivojlantirish" ustuvor yo'nalishi doirasida bajarilgan.

Muammoning o'rganilganlik darajasi. Ko'plab dunyo olimlarining tadqiqotlari NLP (Natural Language Processing) sohasidagi bir qator muammolarni tadqiq etganligi bilan ahamiyatli. Dunyo kompyuter tilshunosligida semantik analiz va analizator borasida bir qator olimlar ilmiy izlanish olib borishgan. A.V.Tuzov, M.V.Mozgovoy, A.V.Sokirko, N.A.Schlayefer, A.V.Mochalova, A.R.Gatiatullin, B.Yergesh, A.Sharipbay, G.Bekmanova, S.Lipniski, Y.A.Kanevskiy, K.K.Boyarskiy, A.Adali kabi olimlarning tadqiqotlarini ana shunday ishlar sifatida qayd etish mumkin.

O'zbek tilini avtomatlashtirish bo'yicha M.Musayev, M.Sharipov, H.Madatov, B.Elov, I.Bakayev, B.Akmuradov, O'.Xamdorov, J. Elov kabi ko'plab olimlarimiz tomonidan taklif qilingan nazariy va amaliy ishlarni ko'rish mumkin. Tilshunos tadqiqotchilar tomonidan tavsiya qilingan lingvistik modellar asosida semantik analizator yaratishga dastlabki qadam qo'yilmoqda.

Tadqiqotning dissertatsiya bajarilgan oliy ta'lim yoki ilmiy-tadqiqot muassasasining ilmiy-tadqiqot ishlari rejalari bilan bog'liqligi. Dissertatsiya Alisher Navoiy nomidagi o'zbek tili va adabiyot universitetining № IL-402104209 raqamli axborot-qidiruv tizimlari (Google, Yandex, Google translate) uchun avtomatik ishlov berish vositasi – o'zbek tilining morfoleksikoni va morfologik analizatori dasturiy vositasini yaratish"(2022-2024) innovatsion loyiha doirasida bajarilgan.

Tadqiqotning maqsadi o'zbek tili semantik analizatori uchun omonim, polisemantik va polifunksional so'zlarni semantik analiz qilish jarayonlarini modellashtirish, ma'lumotlar bazasi va tuzilmalarini yaratish hamda axborot tizimi ishlab chiqishdan iborat.

Tadqiqotning vazifalari:

tabiiy tilga ishlov berish sohasidagi axborotlashtirish jarayonlari, ilmiy tadqiqot ishlari va joriy etilgan axborot tizimlarining xususiyatlarini o'rganish va tahlil qilish;

o'zbek tilidagi gaplarni semantik analiz qilish axborot tizimining konseptual modelini ishlab chiqish;

ma'lumotlar bazasini normallashtirish qoidalari asosida o'zbek tilidagi omonim, polisemantik va polifunksional so'zlarni semantik farqlovchi axborot tizimining axborot modeli hamda ma'lumotlar bazasini ER modelini ishlab chiqish;

omonim, polisemantik va polifunksional so'zlarni xususiyatlaridan kelib chiqib ularni semantik farqlovchi qoidalar ishlab chiqish va usullarni aniqlash;

o'zbek tili gaplarni semantik farqlovchi axborot tizimi uchun omonim, polisemantik, polifunksional so'zlarni farqlovchi biznes-jarayonlarining BPM (Business Process Model) modellarini ishlab chiqish;

semantik analizator axborot tizimining arxitekturasini ishlab chiqish hamda funksional jarayonlarini modellashtirish;

o'zbek tilidagi gaplarni semantik analiz qilish axborot tizimini ishlab chiqish.

Tadqiqot obyekti sifatida o'zbek tilidagi omonim, polifunksional va polisemantik so'zlarni semantik analiz qilish jarayonlari va tizimlari olingan.

Tadqiqot predmetini omonim, polifunksional va polisemantik so'zlarni farqlovchi usullar, modellar hamda jarayonlar tashkil etadi.

Tadqiqot usullari. Tadqiqotda axborot tizimlari va relyatsion ma'lumotlar bazasini loyihalash, strukturaviy va funksional modellashtirish, biznes jarayonlarni modellashtirish asosida axborot tizimini ishlab chiqish usullaridan foydalanilgan.

Tadqiqotning ilmiy yangiligi quyidagilardan iborat:

o'zbek tilidagi omonim, polifunksional va polisemantik so'zlarni kontekstda farqlash usullari aniqlandi hamda semantik analizator axborot tizimining konseptual modeli ishlab chiqilgan;

ma'lumotlar bazasining asosiy atributlari yordamida ma'lumotlarni jamlash, ishlov berish uchun ma'lumotlarning relyatsion modeli va normallashtirish qoidalari asosida o'zbek tilidagi omonim, polifunksional va polisemantik so'zlarni semantik farqlovchi axborot tizimining ma'lumotlar bazasining ER modeli ishlab chiqilgan;

o'zbek tilidagi omonim, polisemantik va polifunksional so'zlarni semantik farqlovchi biznes-jarayonlarining BPM modellari ishlab chiqilgan va jarayonlari lingvistik ma'lumotlar bilan ta'minlangan;

statistik ma'lumotlarga asoslangan holda gap tarkibidagi omonimiya, polisemiya va polifunksionallikni aniqlash, qaror qabul qilishga ko'maklashuvchi ma'lumotlarni jamlash, ishlov berish va taqdim etish axborot tizimi ishlab chiqilgan.

Tadqiqotning amaliy natijalari quyidagilardan iborat:

semantik analizator axborot tizimining konseptual modeli, funksional jarayonlarining BPM modellari ishlab chiqilgan;

omonim, polisemantik va polifunksional so'zlarni farqlovchi axborot tizimining ma'lumotlar bazasining ER modeli ishlab chiqilgan;

omonimiya, polisemiya va polifunksionallikni farqlovchi axborot tizimining axborot modellari ishlab chiqilgan;

korpusda omonimiya, polifunksional va ko'p ma'noli so'zlarni farqlash, ma'lumotlarni kiritish, o'zgartirish, o'chirish va foydalanuvchilarga taqdim etish uchun qulay interfeys ishlab chiqilgan.

Tadqiqot natijalarining ishonchliligi o'rganilgan materiallarning o'zbek tili

tabiatidan kelib chiqqan holda xulosalar qilishga yordam berganligi, ularning asosli ekanligi o'zbek tili semantik analizatorining modellarini, jarayonlarini loyihalashtirishda amalda isbotlangan manbalar, o'zbek tilining lingvistik asoslariga tayanilganligi bilan izohlanadi.

Tadqiqot natijalarining ilmiy va amaliy ahamiyati. Tadqiqot o'zbek tili semantik analizatori ishlab chiqishning nazariy asoslari va uning jarayonlari hamda modellarini ishlab chiqishda soha mutaxassislariga tadqiqotlar olib borishda nazariy va amaliy ahamiyat kasb etadi. O'zbek tili milliy korpusining takomillashishi hamda mukammallik darajasini oshirish imkoniyatlarini beradi.

Tadqiqot natijalarining joriy qilinishi.

Semantik analizatorning axborot-qidiruv tizimidagi ahamiyati, axborot qidiruv tizimida omonim, polifunksional hamda ko'p ma'noli so'zlarni farqlashning matematik modellari, o'zbek tili semantik analizatori uchun omonim, polifunksional va ko'p ma'noli so'zlarni farqlashni amalga oshiruvchi matematik modellar, biznes-jarayon va axborot tizimidan:

tadqiqot davomida ishlab chiqilgan matematik modellar va algoritmlar majmuasini joriy qilish va sinovdan o'tkazish natijasida o'zbek tilida mavjud bo'lgan jami 1065 dan ortiq omonim so'zlarni farqlovchi jami 19 ta matematik model, polisemantik so'zlarni farqlovchi matematik mantiqqa asoslangan 1 ta model va polifunksional so'zlarni semantik farqlovchi jami 5 ta matematik model ishlab chiqilgan;

o'zbek tilidagi omonim va polisemantik so'zlarni semantik farqlash uchun 200 dan ortiq so'z guruhlari, 50 dan ortiq qo'shimchalar va qo'shimchalar kombinatsiyalari guruhlari ishlab chiqilgan;

omonimiyani aniqlashda korpus ma'lumotlari asosida statistik usullardan foydalanish metodlari asoslangan;

omonimiya, polisemiya va polifunksionallikni aniqlashni amalga oshiruvchi axborot tizimning ma'lumotlar bazasining ERD modeli ishlab chiqilgan;

tadqiqot davomida o'zbek tilidagi gaplarni semantik analiz qiluvchi axborot tizimini ishlab chiqishda qo'llanilgan "Qoidalarga asoslangan usul", "Statistik ma'lumotlarga asoslangan usul" va "Mashinali o'qitishga asoslangan usul" lar yordamida omonimiya masalasini yechishga erishilgan (Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universitetining 2022 yil 22 noyabrda dalolatnomasi). Natijada ishlab chiqilgan matematik modellar va biznes-jarayonlar majmuasini joriy qilish va sinovdan o'tkazish natijasida o'zbek tilida mavjud bo'lgan jami 3000 ga yaqin omonim so'zlarni farqlovchi usullar aniqlangan, 19 ta matematik model, polisemantik so'zlarni farqlovchi matematik mantiqqa asoslangan 1 ta model va polifunksional so'zlarni semantik farqlovchi jami 5 ta matematik model ishlab chiqilgan; tadqiqot davomida ishlab chiqilgan axborot tizimida "Qoidalarga asoslangan usul", "Chastotali usul", "Naive Bayes klassifikatori" hamda Yashirin Markov modellaridan foydalanilgan va ishlab chiqilgan biznes-jarayonlar asosida omonimiyani bartaraf etish aniqligi 84 foizni, polisemiyaning farqlash aniqligi 73 foizni va polifunksionallikni farqlash aniqligi 74.5 foizni tashkil etgan.

tadqiqot natijalaridan Xorazm viloyati Urganch shahrida joylashgan "Mustafo Software MCHJ" dasturlash firmasi dasturchilari tomonidan ishlab chiqilgan

“MyDoctor consultant” dasturini yaratishda foydalanildi (Mustafo Software MCHJ” direktorining 2023-yil 1-fevraldagi dalolatnomasi). Tadqiqot natijalaridan foydalanish natijasida inson a’zolari orasidagi ko‘p ma’nolilikni aniqlashga erishildi. Berilgan xulosalardagi ma’nolarni aniqlashda semantik o‘xshashliklarni aniqlash natijalaridan foydalanildi. Bu “MyDoctor consultant” dasturining samaradorligini 1.2 barobarga oshirishga xizmat qildi.

Tadqiqot natijalarining aprobatsiyasi. Mazkur tadqiqot natijalari 8 ta xalqaro, 5 ta respublika anjumanida muhokamadan o‘tkazilgan.

Tadqiqot natijalarining e’lon qilinishi. Dissertasiya mavzusi bo‘yicha 30 ta ilmiy ish nashr ettirilgan, jumladan, O‘zbekiston Respublikasi Vazirlar Mahkamasi huzuridagi Oliy attestasiya komissiyasining doktorlik dissertatsiyalari asosiy ilmiy natijalarini chop etish uchun tavsiya etilgan ilmiy nashrlarda 8 ta ilmiy maqola, ulardan 4 tasi xorijiy (2 ta Scopus bazada indekslangan) jurnallarda chop qilingan va 1 ta monografiya nashr etilgan.

Dissertatsiyaning hajmi va tuzilishi. Dissertasiya kirish, uchta bob, xulosa, foydalanilgan adabiyotlar ro‘yxati va ilovadan iborat. Hajmi 118 sahifani tashkil etadi.

DISSERTATSIYANING ASOSIY MAZMUNI

Kirish qismida mavzuning dolzarbligi va zarurati asoslangan, tadqiqotning respublika fan va texnologiyalari rivojlanishining ustuvor yo‘nalishlariga bog‘liqligi ko‘rsatilgan, maqsad va vazifalari belgilangan, obyekt va predmeti tavsiflangan, ilmiy yangiligi va amaliy natijalari bayon qilingan, natijalarning ilmiy va amaliy ahamiyati ochib berilgan, joriylanishi, aprobatsiyasi, nashr etilgan ishlar va dissertatsiya tuzilishi bo‘yicha ma’lumotlar keltirilgan.

Dissertatsiyaning birinchi bobi **“Gaplarni analiz qilish usullari va tizimlari”** deb nomlangan. Bobning *“Tabiiy tillarda gaplarni tahlil qilish tizimlari”* nomli birinchi bo‘limida gap analizatorlarining umumiy masalalari, ularning turlari, dunyo va o‘zbek kompyuter tilshunosligida analizator tadqiqi masalalari tahlil etilgan. Bo‘limda lingvistik analizatorning turlari *morofologik, sintaktik, semantik analizatorlar* va ularning vazifalari haqida umumiy ma’lumotlar berilgan.

Bobning *“Gaplarni semantik analiz qilishning model va metodlari”* deb nomlangan ikkinchi bo‘limida matnlarni semantik analiz qiluvchi model va metodlar haqida fikr yuritilgan. Semantik analizni amalga oshiruvchi axborot tizimi turli tillarda turlicha model va algoritmlarga asoslanadi. Bobda omonim, polisemantik va polifunksional so‘zlarni semantik analiz qiluvchi tizimni ishlab chiqishdagi zarur bo‘lgan yondashuvlar haqida ma’lumot berilgan. Omonimiyani aniqlashning mavjud yondashuvlari lokal va global sintaktik tahlil, semantik lug‘at va ehtimolliklarga asoslanadi. Grammatik va statistik hodisalardan foydalangan holda korpusdagi so‘zlarning belgilari asosida omonimiya hodisasi aniqlanadi. Omonimiyani muvaffaqiyatli bo‘lishi uchun aniq natijalar kerak. Bundan tashqari ma’lumotlar hajmining kattaligi ularga ishlov berishni sekinlashtiradi, bu esa **kombinator portlash** deb ataladi. Jahon kompyuter lingvistikasida gaplarni semantik tahlil qilishda asosan 3 ta yondashuvdan foydalanilgan: *Qoidalarga asoslangan yondashuv*;

Statistik ma'lumotlarga asoslangan yondashuv; Mashinali o'rganishga asoslangan yondashuv.

Qoidalarga asoslangan yondashuv – soʻzlarni til grammatikasining qoidalari yordamida semantik analiz qilish nazarda tutilgan. Oʻzbek tilida qator qoidalar ishlab chiqilgan. Dissertatsiya ushbu qoidalarga asoslangan holda baʼzi turdagi omonim soʻzlarni farqlovchi matematik modellar ishlab chiqilgan. Ushbu yondashuvning mohiyati shundan iboratki, baʼzi hollarda kontekst tahlili gapning bir qismining sintaktik tuzilishini va u yordamida soʻzshakllarini tushunishga yordam beradi.

Statistik ma'lumotlarga asoslangan yondashuv – masalani yechishda soʻzlarning grammatik parametrlarini tasniflash orqali foydalaniladi. Bu parametrlar turli tabiiy tillarda turlicha tanlab olinadi. Tasniflash parametrlari sifatida morfologik tahlil, lemma, oʻzak, oʻzak + qoʻshimcha kabi parametrlar boʻlishi mumkin. Soʻzning semantik azalizining aniqlik darajasi yuqori boʻlmasa, tasniflanish parametrlarini kengaytirishga zarurat tugʻiladi. Ushbu yondashuvning asosiy vazifasi kontekst tarkibini n-grammlarga ajratish, yaʼni kontekstdagi kirish soʻzining birikuvchilari aniqlab, baholash metodlari yordamida baholanadi. Baʼzi statistik usullarda soʻz va uning qoʻshimchalari orqali qaror qabul qilinsa, baʼzilarida soʻzning kontekstdagi semantik valentliklari(birikuvchilari) yordamida xulosa qilinadi.

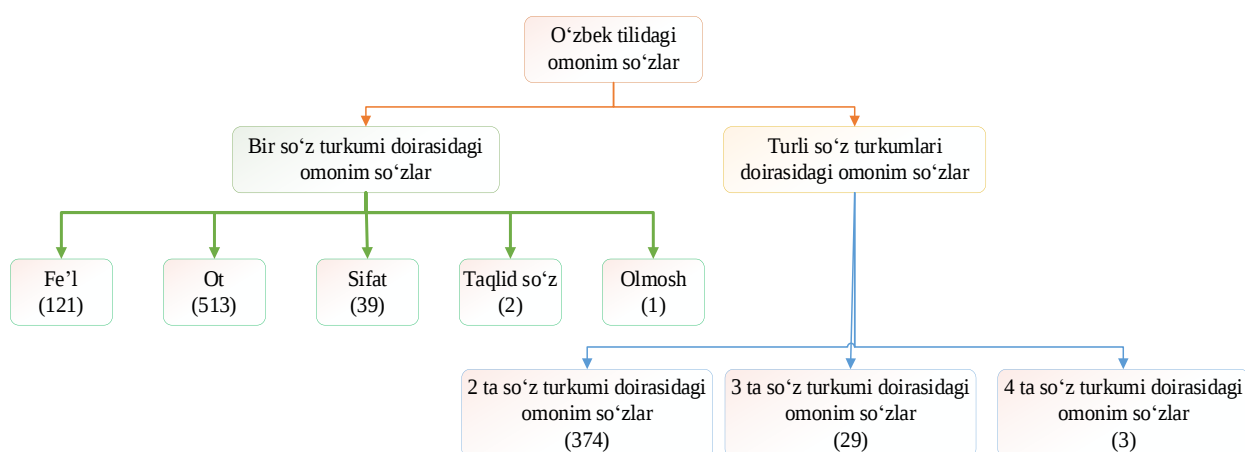
Mashinali o'rganishga asoslangan yondashuv –statistik usullarga qoʻshimcha boʻlib, jahon olimlari hozirda omonimiyani aniqlashda keng foydalanmoqdalar. Mashinali oʻrganishga asoslangan yondashuv yordamida oʻzbek tilidagi bir soʻz turkumi doirasidagi omonimiyani aniqlash mumkin degan xulosaga kelindi. Koʻplab xorijiy tajribalar shuni koʻrsatadiki, ushbu yondashuvdan foydalanish uchun maʼnosi oldindan aniqlangan, belgilangan matnlar kerak boʻladi. Bundan kelib chiqadiki, oʻzbek tili milliy korpusi maʼlumotlar bazasidagi katta hajmdagi matnlarni semantik teglash, maʼnolarini belgilab olish dolzarb masaladir.

Bobning "*Oʻzbek tili semantik analizatorida matnlarni semantik teglash usullari*" nomli uchinchi boʻlimida oʻzbek tili semantik analizatorini yaratishi jarayonida foydalaniladigan matnlarni teglash masalasi tavsiflangan. Tabiiy tillarga ishlov berish jarayoni (NLP-Natural languages processing) ning muammolarini ketma-ket modellashtirish zamon talabidir. Matnda omonim va polifunksional soʻzlarning aniqlanishida bevosita soʻz turkumlarini teglash muammosiga duch kelish mumkin. Masalan, oʻzbek tilidagi ikki soʻz turkumi doirasidagi omonim soʻzlarning maʼnolarini kontekstda aniqlash jarayonida *Qoidalarga asoslangan teglash usuli ham, Stoxastik teglash usuli* ham muhim ahamiyat kasb etadi.

Qoidalarga asoslangan teglash. Soʻz turkumlarini avtomatik teglash odatda qoidalarga asoslangan yondashuvlar nomaʼlum yoki noaniq soʻzlarga teglar belgilash uchun kontekstli maʼlumotlardan foydalanadi.

Stoxastik teglash usuli. "Stochastic tagging" ST atamasi tegi muammosiga turli xil yondashuvlarni nazarda tutishi mumkin. Baʼzi manbalarda statistik metod deb ham qoʻllaniladi. Har qanday holatda chastota yoki ehtimollikni oʻz ichiga olgan har qanday model *toʻgʻri stoxastik model* deb belgilanadi. Eng oddiy stoxastik teglash soʻzlarni faqatgina soʻzning maʼlum bir teg bilan paydo boʻlish ehtimoli asosida ajratib turadi.

Dissertatsiyaning **II bobi “Soʻzlarni semantik analiz qilish usullari”** deb nomlangan. Bobning “*Oʻzbek tilida omonimiyani bartaraf etish usullari*” nomli birinchi boʻlimida oʻzbek tilidagi omonim soʻzlarni semantik farqlovchi matematik modellar oʻrganilgan. Oʻzbek tilidagi omonim soʻzlarni semantik farqlashda ikki guruhga: *bir soʻz turkumi doirasidagi va turli soʻz turkumi doirasidagi omonim* soʻzlarga ajratilgan. Har bir guruhdagi omonim soʻzlarni farqlovchi matematik mulohazalar ishlab chiqildi. Omonim soʻzlarni soʻz turkumlari doirasida 1-rasmdagi kabi klassifikatsiya qilingan. 2-rasmda keltirilgan iyerarxiyadan tasvirlanganidek bir soʻz turkumi, yaʼni feʼl, ot, sifat, undov va olmosh soʻz turkumi doirasidagi omonim soʻzlar kuzatilgani haqida Sh.Gulyamovning tadqiqotlarida taʼkidlangan.

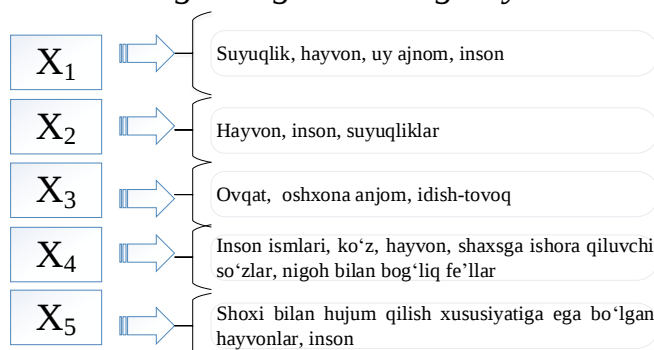


1-rasm. Oʻzbek tilidagi omonim soʻzlar iyerarxiyasi

Bir soʻz turkumi doirasidagi omonimiyani aniqlashni feʼl soʻz turkumida 5 xil maʼnoga anglatuvchi *suzmoq* feʼli misolida keltiramiz. *Suzmoq* feʼlining maʼnolari 2-rasmda keltirilgan. **Suz** soʻzi kontekstda turli maʼnoni anglatada, bir xil qoʻshimcha bilan birikib kelgan holatlar koʻp uchraydi. Mazkur masalani yechishda Mashinali oʻrganishga asoslangan usuldan foydalanish tavsiya qilingan, chunki bir soʻz turkumi doirasidagi omonimlikni farqlashda hech qanday fiksirlangan qoida mavjud emas. Agar boʻlsa ham umumiy emas. Qoidalarga asoslangan yondoshuv yordamida ham bu masalaga yechim topib boʻlmazligi ham asoslangan. Demak, bir soʻz turkumi doirasidagi omonimiyani aniqlash *mashinali oʻrganishga asoslangan yondashuvga*

X ₁	≡	Suyuqlikni filtrlamoq
X ₂	≡	Suvda harakatlanmoq
X ₃	≡	Ovqatni idishga solmoq
X ₄	≡	Noz-karashma qilmoq
X ₅	≡	Shohi bilan hujum qilmoq

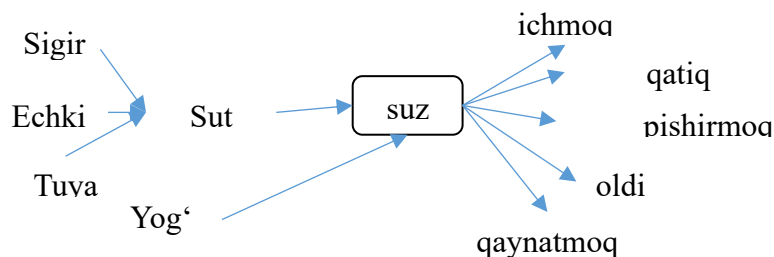
2-rasm. Suz soʻzining maʼnolari



3-rasm. Suz soʻzining semantik birikuvchilarining soʻz guruhlari

murojaat qilishni taqozo etadi. Tadqiqot ishining bajarilishi davomida neyron tarmoqlarga murojaat qilinmagan. Ammo bir soʻz turkumi doirasidagi omonimiyani aniqlashni boshqa bir usulini, yaʼni neyron tarmoqlardan foydalanishda zarur

bo'ladigan so'z guruhlaridan foydalangan holda yechish taklif qilingan. Darhaqiqat, bu usul yordamida to'liq maqsadga erishilmaydi albatta, lekin ma'lum darajada vazifa bajariladi. Omonim so'zning semantik valentliklari yordamida aniqlash jahon tajribasida *semantik konkordanse* usuli deb nomlanadi. *Suzmoq* fe'lining ma'nolarini x_i bilan belgilangan (2-rasm). Demak, 5 ta yashirin holat mavjud, ya'ni yashirinligi shundaki, berilgan gapda uchragan *suzmoq* fe'lining aynan qaysi ma'noni anglatayotganligi noma'lum. Berilgan gapdagi *suzmoq* fe'lining aynan qaysi ma'noda kelganligini aniqlash uchun kontekstdan foydalaniladi. Buning uchun omonim so'zning semantik valentliklari (birikuvchilari) oldindan belgilab qo'yilgan bo'lishi kerak. *Suz* omonim so'zning semantik birikuvchilari guruhlar bilan birikkanida 4-rasmdagi kabi ma'nolarni anglatishi aniqlandi. Ushbu rasmda tasvirlangan har bir so'z o'zining semantik iyerarxiyasiga ega. Demak kontekst orqali so'z ma'nosini aniqlash uchun o'zbek tilida mavjud bo'lgan so'zlarni semantik guruhlariga ajratib chiqish kerak degan xulosaga kelish mumkin. 4-rasmdagi *suzmoq* fe'lining birikuvchilari *sut emizuvchi hayvonlar, filtrlanish xususiyatiga ega bo'lgan suyuqliklar, suyuqlik bilan bog'liq bo'lgan harakat va holat fe'llari, oshxona anjomlari* kabi so'z guruhlariga mansub so'zlardan iborat bo'lganda suyuqlikni filtrlamoq manosi anglatishi mumkin.



4-rasm. Suzmoq fe'lining semantik valentliklar to'plami

Shu yerda savol paydo bo'ladi. *Bir so'z turkumi doirasida omonimlik hosil qiluvchi so'zlar ko'p ularning har biri uchun shunday so'z guruhlarini aniqlab chiqiladimi?* Ushbu savolga javob topish uchun ko'plab ilmiy izlanishlarni tahlil qilindi, jahon tajribasi o'rganildi. Nihoyat, buning boshqa yechimi yo'q degan xulosaga kelindi. Omonim so'zlarni o'zaro semantik guruhlariga ajratish kerak degan fikr ham paydo bo'ldi, ammo bu fikr ham o'zini oqlamadi. Bir so'z turkumidagi omonim so'zlarni har birining semantik valentliklarini aniqlash va ularni semantik guruhlariga ajratishda katta hajmdagi ma'lumotlar, asarlar, maqolalar kerak bo'ladi. Bu esa bizda mavjud, ya'ni O'zbek tili milliy korpusi¹ ma'lumotlar bazasidan 2 milliarddan ortiq matnlar joy olgan. So'z guruhlarini so'zlar bilan to'ldirish, omonim so'zlarni birikuvchilarni ajratib olish va belgilash kabi ishlar korpus ma'lumotlari yordamida mutaxassislar tomonidan amalga oshirilishi ancha qulay ekanligi haqida aytib o'tilgan. Bo'limning *Turli so'z turkumlari doirasidagi omonimiyani bartaraf etish usullari* deb nomlangan faslida esa o'zbek tilidagi turli so'z turkumlari doirasidagi omonim so'zlarni farqlashda *qoidalarga asoslangan va statistik ma'lumotlarga asoslangan* yondashuvlardan foydalanish haqida yoritib berilgan.

¹ <http://uzschoolcorpara.uz/uz>

O'zbek tilidagi omonim so'zlar turli so'z turkumlari doirasida omonimlik hosil qilishiga ko'ra uchta (2,3,4 so'z turkumlari doirasida) guruhga ajratilgan va so'z turkumlari guruhlari ham aniqlangan. Turli so'z turkumlari doirasidagi omonimiyani farqlashda qoidalarga asoslangan usuldan foydalanilgan. Misol tariqasida uchta so'z turkumi (ot v son v fe'l) doirasida omonimlik hosil qiluvchi so'zlarni farqlovchi model keltirilgan:

- 1) **Uch:** Biror narsaning boshlanish yoki tugash qismi. (ot)
- 2) **Uch:** Ikki bilan to'rt orasidagi butun son. (Son)
- 3) **Uch:** 1. Havo bo'ylab harakatlanmoq. (fe'l)
2. Biror qirrasi shart uzilib tushmoq.
3. Beixtiyor berilmoq, laqillamoq

Kuzatuvlar natijasida quyidagi qoidalarga yig'ildi:

- **uch**+*-i/-iga/-ni/-da* ⊕ *bor/ulamoq/boshlamoq/turmoq/...* (ot)
- **uch**+*-inchi/-nchi/-ta/-tacha/-tadan/-larcha/* ⊕ *oy nomlari/dars/uy/...* (son)
- **uch**+*-ov/-ovlon/..* ⊕ *gaplashmoq/ko'rishmoq/kelmoq/ketmoq/...* (son)
- **uch**+*dan* ⊕ *bir/ikki/uch/to'rt/...* (son)
- *Balandga/osmonga/...* ⊕ **uch** (fe'l)
- *Toshkentdan/shahardan...* ⊕ **uch** (fe'l)

Yuqoridagi morfologik xususiyatlarni inobatga olib son so'z turkumidagi omonim farqlovchi modelni quyidagicha keltirish mumkin.

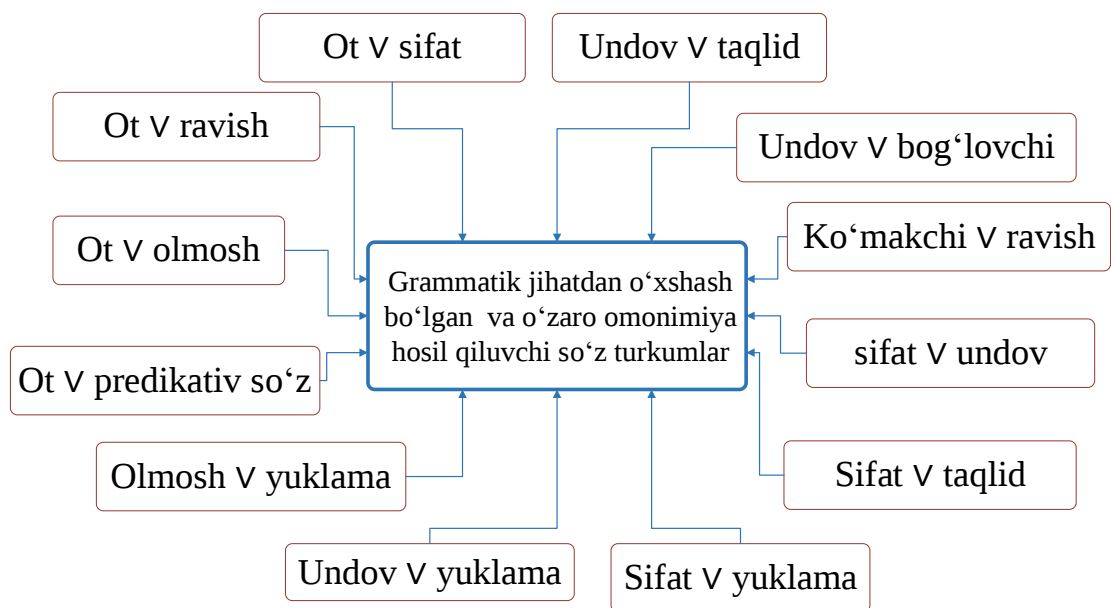
$$H_{num,N,V} = \begin{cases} H^{num} + \downarrow aff_l^{ord} \oplus \downarrow (W_k^N + \downarrow aff_l^{ord}) \oplus N_j^c \\ \downarrow W_i^{dec} \oplus H^{num} + \downarrow aff_l^{ord} \oplus N_j^c \\ H^{num} + aff_j^{agg} \oplus V \\ H^{num} + dan \oplus W_n^{num} \\ N + aff_i^N \\ V + aff_j^V \end{cases} \quad (1)$$

Bunda,

aff_j^{agg} –jamlovchi qo'shimchalar to'plami, $aff_j^{agg} = \{-ov, -ovlon, -ala, -ovi, -oviga, -ovloni, \dots\}$; W_n^{num} –sonlarning nomlaridan iborat so'zlar to'plami, $W_n^{num} = \{bir, ikki, uch, \dots\}$

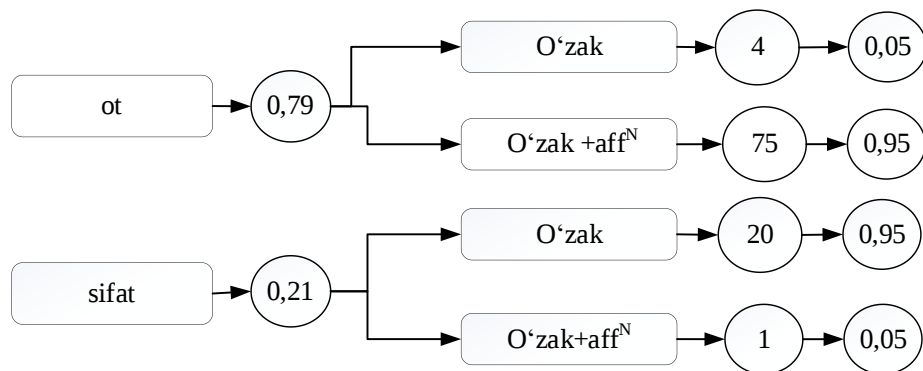
Turli so'z turkumlari orasida shunday so'z turkumlari borki (5-rasm), grammatik jihatdan o'xshash, ular orasidagi omonimiyani aniqlashda qat'iy qoidalar ish bermaydi. Shuning uchun statistik ma'lumotlarga asoslangan usuldan foydalangan ma'qul. Chastotali usul ana shunday statistik usullardan biridir. *Ot yoki sifat* so'z turkumi doirasidagi omonimiyani Chastotali usul yordamida aniqlashni ko'rib chiqamiz. Chastotali usuldan foydalanish uchun so'z turkumlarini farqlovchi tasniflash parametrlari ajratib olishni talab etadi. *Ot V sifat* so'z turkumi doirasidagi omonim so'zlar quyidagi parametrlar bo'yicha tasniflanadi:

1. So'z turkumi;
2. O'zak va lemma;
3. Faqat lemma
4. Faqat o'zak



5-rasm. Grammatik jihatdan o'xshash bo'lgan va o'zaro omonimiya hosil qiluvchi so'z turkumlari.

Uzschoolcorpora.uz sayti orqali sifat V ot so'z turkumi doirasida omonimlik hosil qiluvchi *alomat* so'zi qidirilganda jami 6823 ta ma'lumot topildi, shulardan 100 tasi tahlil qilinganda 79 tasi *ot* so'z turkumi, 21 tasi *sifat* so'z turkumi ekanligi aniqlandi. Aniqlangan ma'lumotlarni o'zak va qo'shimchalarga ajratilib tahlil qilinganida quyidagi ehtimolliklar hisoblandi.



6-rasm. Alomat so'zining korpusda uchrash holatlarining ehtimolliklari

Tasniflash parametrlar asosida hisoblangan ehtimolliklar shuni ko'rsatadiki, *ot* yoki *sifat* so'z turkumlari doirasida omonimiya hosil qiluvchi so'zlarni ot so'z turkumiga oid deya baholanadi, qachonki tarkibida ot so'z turkumiga xos qo'shimcha bilan birikkan bo'lsa. Undan tashqari turli so'z turkumlari doirasidagi omonimiyani aniqlashda Naïve Bayes usulidan foydalanish haqida ham yoritib berilgan.

Bobning "*O'zbek tilidagi polisemantik so'zlarni aniqlash usullari*" deb nomlangan uchinchi bo'limida polisemiyaning farqlashning usullari haqida so'z yuritiladi. Bu bo'limda *ko'z* so'zi misolida polisemiyaning aniqlashning matematik modeli ishlab chiqilgan. Polisemiyaning aniqlashda segmentatsiya usulidan foydalanish asoslangan.

Anor xolaning ko'zi tekkan aslida, aks holda o'qishga kirgan bo'lardi.

Namunada berilgan *ko'z ko'p ma'noli so'zi xola so'zi bilan birikkan*, lekin *ko'z* so'zidan keyin *tekkan* so'zining uchrashi *ko'chma ma'noga* olib keladi. Undan kelib chiqadiki, *ko'p ma'noli so'zning o'ng va chap birikuvchilari o'z va ko'chma ma'no* hosil qilishiga *ko'ra ham guruhlarga ajratilishi lozim*. *Ko'z so'zining o'z va ko'chma ma'nolarini* hosil qiluvchi birikuvchilari tasvirlangan ierarxiyalarda ko'rsatilgani kabi so'zlarni guruhlarga ajratish kerak. So'zlar qancha kichik-kichik guruhlarga ajratilsa *ko'p ma'noli so'zlarning semantik tahlili shuncha aniq bo'ladi*. Xorijiy tajribalarga ko'ra *ko'p ma'noli so'zlarni semantik tahlil qilishda statistik metodlardan foydalanish eng ma'qul yechim hisoblanadi*.

Ko'z so'zining birikuvchilarini o'z va ko'chma ma'no anglatishiga ko'ra guruhlarga ajratib olinadi.

$w_i^b - P$ polisemantik so'zning chap tomonidan birikib asl ma'nosini hosil qiluvchi so'zlar to'plami;

$w_j^a - P$ polisemantik so'zning o'ng tomonidan birikib asl ma'nosini hosil qiluvchi so'zlar to'plami.

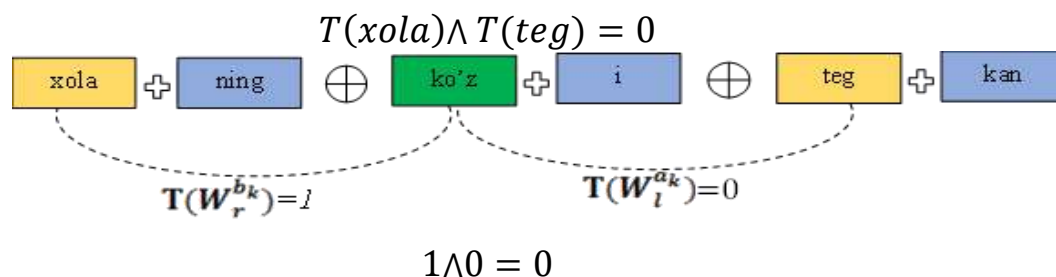
Ajratilgan birikuvchilar to'plamlarini mavjud deb faraz qilib, quyidagi matematik modeli keltirilgan.

$$w_r^{b_k} + \downarrow aff^{s_l} \oplus P_k + \downarrow aff^{s_l} \oplus w_l^{a_k} + \downarrow aff^{s_l} = \begin{cases} P_k^s, T(w_r^{b_k}) \wedge T(w_l^{a_k}) = 0 \\ P_k^p, T(w_r^{b_k}) \wedge T(w_l^{a_k}) = 1 \end{cases} \quad (2)$$

(2) model yordamida quyidagi jumladagi *ko'z* so'zini tahlil qilishni ko'rib chiqilgan.

Anor xolaning **ko'zi tekkan** aslida, aks holda o'qishga kirgan bo'lardi

Jumladagi *ko'z* so'zi o'ng tomonidan *xola* (*ko'rish xususiyatiga ega*) so'zi bilan birikkan, demak $xola \in \{inson\}$ guruhiga mansub so'z va $T(xola)=1$ - *xola* so'zi bilan *ko'z* so'zi birikib bu *ko'p ma'noli so'z o'z* (*ko'rish organi*) ma'nosini anglatadi; Chap tomonidagi *teg* so'zi *ko'zning funksiyasiga tegishli* emas va $T(teg)=0$. Demak,



keltirilgan gapdagi *ko'z* so'zi uchun (3) matematik model yordamida yuqori natijaga erishish uchun o'zbek tilida mavjud so'zlar semantik guruhlarga ajratilishi lozim.

Bobning uchinchi "*Kontekstda polifunksional so'zlarni semantik farqlovchi modellar*" bo'limida o'zbek tilidagi polifunksional so'zlarning klassifikatsiyasi keltirilgan. Polifunksional so'zlar ham omonimlar kabi morfologik xususiyatlari asosida farqlanishi mumkin. Ammo bu usulning o'zi barcha polifunksional so'zlarni farqlab, model tuzish uchun yetarli emas. Chunki ba'zi so'z turkumi kontekstda aniqlanadi. Garchi shunday bo'lsa-da, polifunksionallikni farqlovchi tizim uchun aniq qonuniyatlar, modellar kerak bo'ladi. Sh.Gulyamova turkumlar kesimida klassifikatsiya qilganida, har bir guruh polifunksional so'zlari uchun lingvistik model

ishlab chiqqan. Ushbu lingvistik modellar asosida matematik modellar ishlab chiqildi. *ot* $\vee$ *sifat* soʻz turkumlari orasidagi polifunksionallikni aniqlovchi matematik model keltirilgan.

$$\begin{aligned} & - \text{Kasal} + -i/ni/ga/lar/da/ini/... \oplus \text{aniqlandi/ko 'rmoq/keldi/...} \\ & - \text{Kasal} \oplus \text{odam/bemor/hayvon/qush/...} \\ Pf_{adj,N} &= \begin{cases} Pf^N + \downarrow aff^N \oplus V \\ Pf^{adj} + \downarrow aff^{deg} \oplus N \end{cases} \end{aligned} \quad (3)$$

(4) yordamida *ot* $\vee$ *sifat* soʻz turkumlari orasidagi polifunksionallikni aniqlash tavsiya qilingan. Xuddi shunday *sifat* $\vee$ *modal*, *sifat* $\vee$ *ravish*, *mustaqil* $\vee$ *ko 'makchi fe'l*, *bog 'lovchi* $\vee$ *modal* soʻz orasidagi polifunksionallikni aniqlovchi matematik modellar ishlab chiqilgan.

$$Pf_{M,Adj} = \begin{cases} Pf^{adj}, \text{ agar } Pf^{M,Adj} + \downarrow aff^{adj} \oplus W^{Pf_{adj}} \\ Pf^M, \text{ agar } W^M \oplus Pf^{M,Adj} \oplus W^M \end{cases} \quad (4)$$

$$Pf_{Adj,Adv} = \begin{cases} Pf^{Adj}, \text{ agar } Pf^{Adj,Adv} + aff^{deg} \oplus N^{Pf_{adj}} \\ Pf^{Adv}, \text{ agar } Pf^{Adj,Adv} \oplus V \end{cases} \quad (5)$$

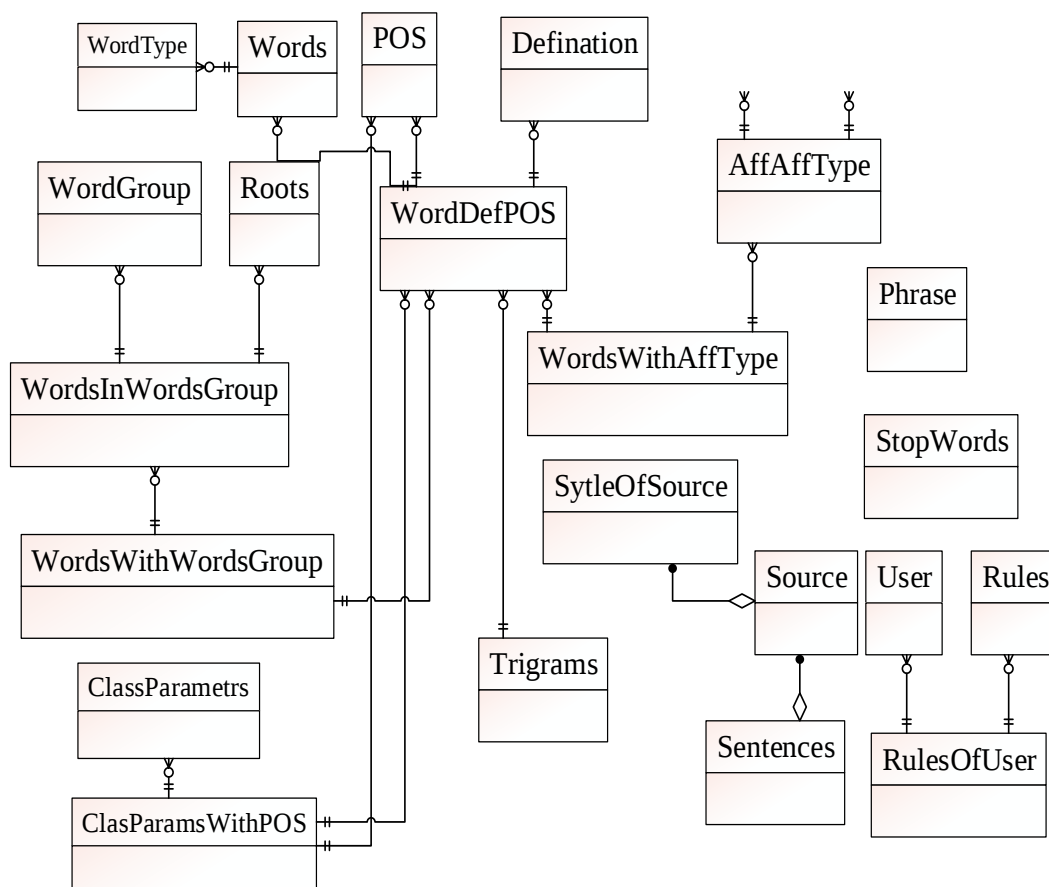
$$Pf_{I,Aux} = \begin{cases} Pf^I, \text{ agar } W^x + aff^{CS} \oplus Pf^{I,L} \\ Pf^{Aux}, \text{ agar } V^{Aux} + aff^{Aux} \oplus Pf^{I,Aux} \end{cases} \quad (6)$$

$$Pf_{Conj,M} = \begin{cases} Pf^{Conj}, \text{ agar } W^{Conj} \oplus Pf^{Conj,Aux} \oplus W^{Conj} \\ Pf^M, \text{ agar } W^M \oplus Pf^{Conj,M} \oplus V \end{cases} \quad (7)$$

Dissertatsiyaning III bobi “O‘zbek tilidagi gaplarni semantik analiz qilish axborot tizimining modellari va jarayonlarini loyihalashtirish” deb nomlangan bo‘lib uchta bo‘limdan iborat. Bobning “O‘zbek tilidagi gaplarni semantik analiz qilish axborot tizimining axborot modelini loyihalashtirish” deb nomlangan birinchi bo‘limida axborot tizimining asosi bo‘lgan axborot modeli (7-rasm) hamda ma’lumotlar bazasining ER modeli ishlab chiqilgan, uning obyektlari, ma’lumotlarning turlari haqida soʻz yuritilgan. O‘zbek tilidagi gaplarni semantik analiz qilish axborot tizimining ma’lumotlar bazasi ER modeli IDEF1X notatsiyasi yordamida Visio platformasida ishlab chiqilgan(9-rasmda).

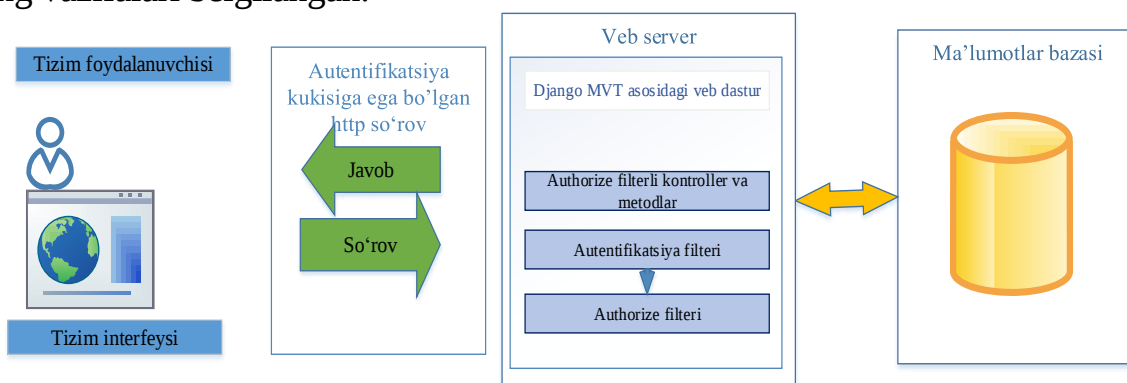
Bobning “O‘zbek tilidagi gaplarni semantik tahlil qilish axborot tizimining funksional jarayonlarini modellashtirish” deb nomlangan ikkinchi bo‘limida semantik analizator axborot tizimining konseptual modeli, o‘zbek tilidagi gap tarkibidagi omonim, polisemantik va polifunksional soʻzlarni semantik tahlil qiluvchi axborot tizimining biznes-jarayonlarining BPM modellari ishlab chiqilgan. 10-rasmda semantik analizator axborot tizimining konseptual modeli keltirilgan. Omonimiyani aniqlashda inson ismlari va leksik birliklarning omonimlik hosil qilishi masalasi ham inobatga olingan. Turli soʻz turkumlari doirasidagi omonimiyani farqlovchi har bir model uchun algoritmlar ishlab chiqilgan.

Bobning “Semantik analizator axborot tizimining imkoniyatlari va tahliliy natijalari” deb nomlangan uchinchi bo‘limida o‘zbek tilidagi gap tarkibidagi omonim, polifunksional va polisemantik soʻzlarni ma’nosini aniqlovchi axborot

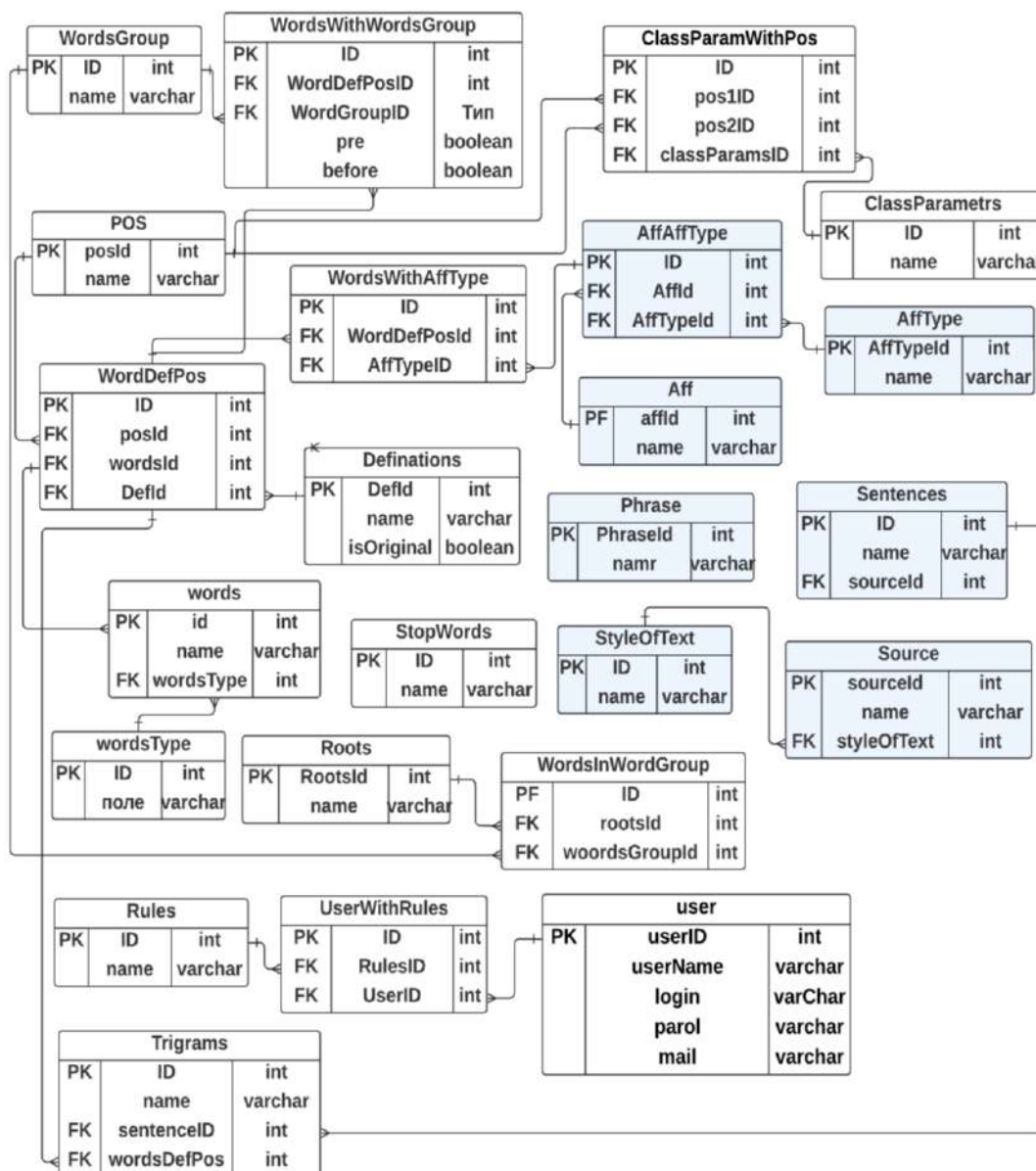


7-rasm. O‘zbek tilidagi gaplarni semantik analiz qilish axborot tizimining axborot modeli

tizimining imkoniyatlari haqida so‘z yuritilgan Undan tashqari semantik analiz va uni o‘rganish haqidagi ma’lumotlar berilgan. Semantik analiz NLP uchun juda muhim, chunki uning jarayonlari so‘zlarning turli ma'nolarini aniqlashga imkon yaratadi. Shu bilan birga, bu jarayonlar mashinaga butun jumlar va matnlarning ma'nosini tushunishga yordam beradi. Semantik analizator axborot tizimi foydalanuvchilari veb brauzer yordamida internet yoki lokal tarmoq orqali murojaat qiladi. Bunday holatda xavfsizlikni ta'minlash uchun veb brauzerdan veb servergacha kelgan trafikning maxfiyligini ta'minlash, foydalanuvchilarni autentifikatsiya qilishni ta'minlash kerak. Tizim foydalanuvchilarining autentifikatsiyasi, identifikatsiyasi va avtorizatsiyasi arxitekturasi 8-rasmda keltirilgan. Foydalanuvchilarning authorizatsiyasini ta'minlash uchun axborot tizimida rollar mexanizmidan foydalanilgan. 12-rasmda rollarning vazifalari belgilangan.



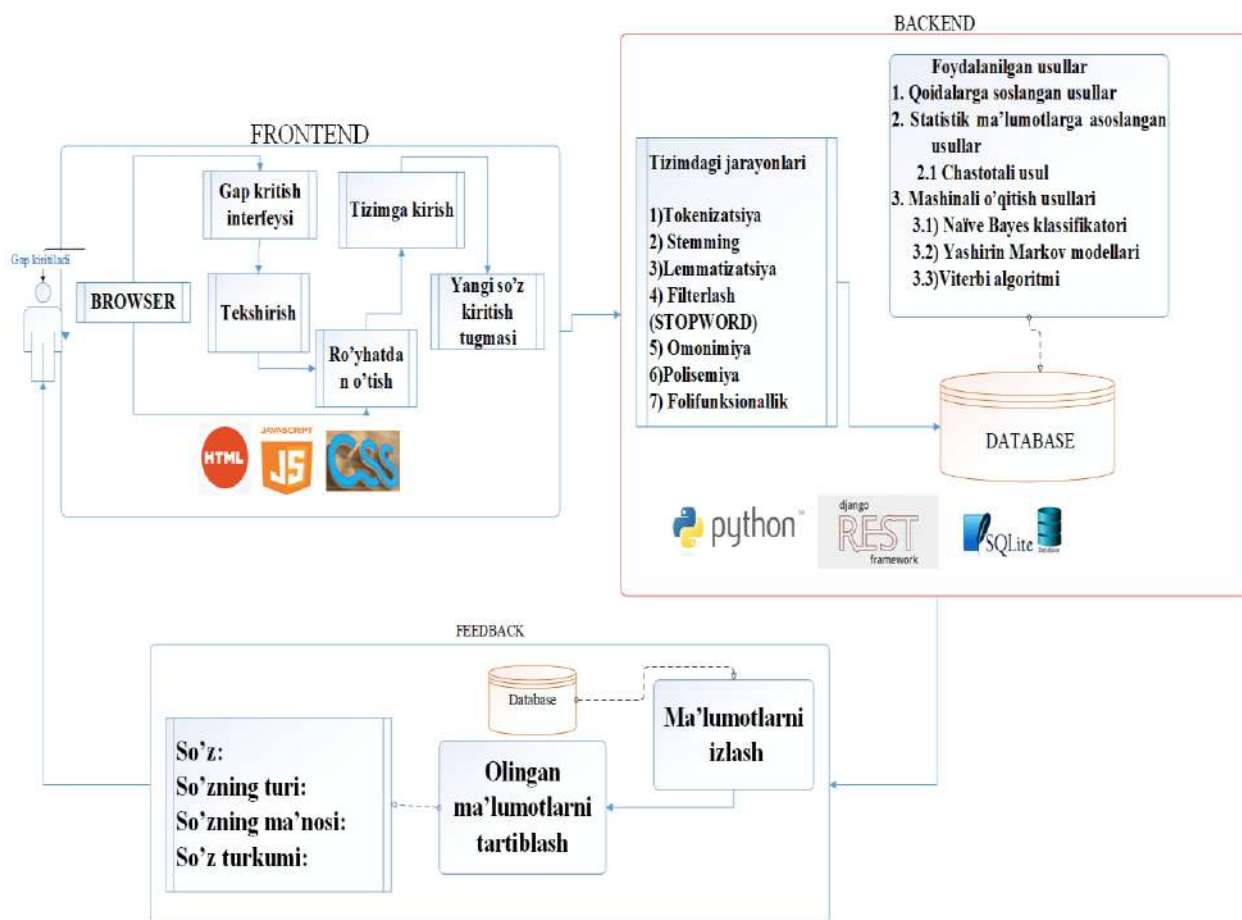
8-rasm. Semantik analizator foydalanuvchilarining autentifikatsiyasi, identifikatsiyasi va avtorizatsiyasi



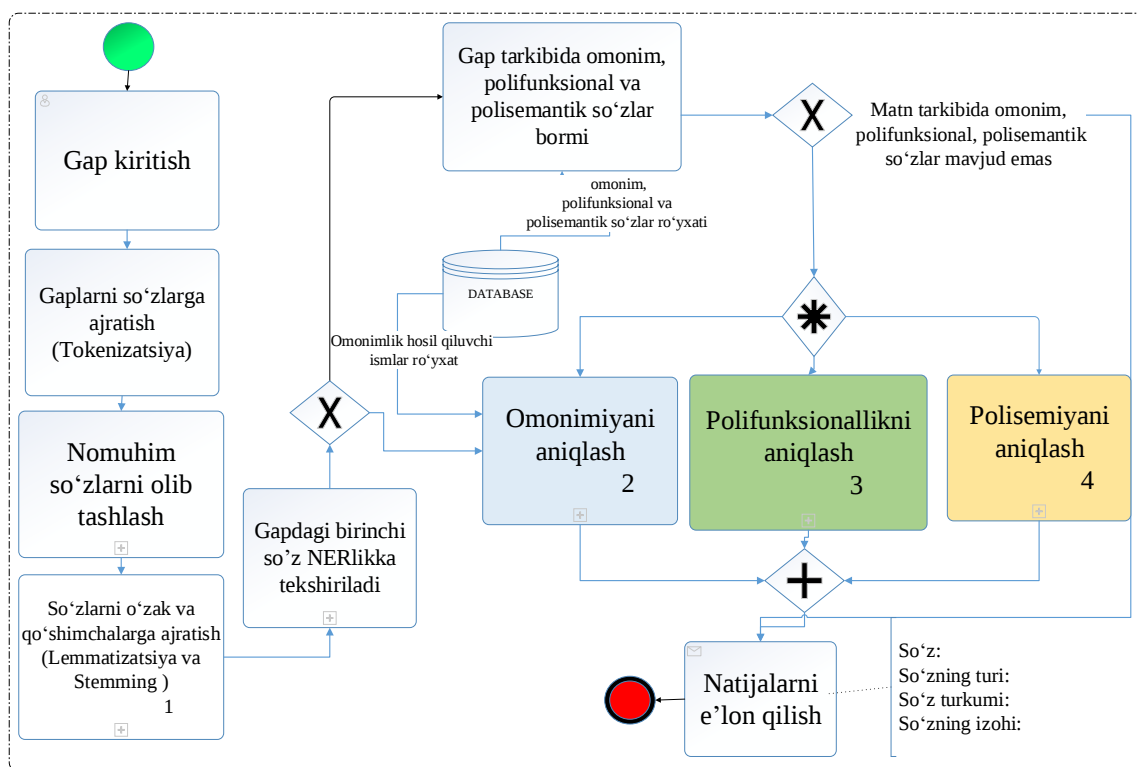
9-rasm. Semantik analizator axborot tizimining ma'lumotlar bazasining ER modeli

Semantik analizator axborot tizimini ishlab chiqishda foydalanilgan usullarning aniqlilik darajasi 1-jadvaldagi kabi asoslandi.

1-jadvalda keltirilgan qiymatlar har bir turdagi so'zlarni aniqlashda ishlab chiqilgan matematik modellar, aniqlangan tasniflash parametrlari hamda so'zlarning semantik guruhlari natijalarida olingan. O'zbek tilidagi gaplarni semantik analiz qilish axborot tizimini ishlab chiqishda asosan qoidalarga asoslangan usul, Chastotali usul, Naïve Bayes klassifikatori hamda Yashirin Markov modellaridan foydalanildi. YMM dan foydalanish uchun mavjud omonim so'zlarning trigrammalaridan iborat o'rtacha 5000 ta so'z birikmalari semantik teglangan holda ma'lumotlar bazasiga joylandi. Omonimiyani aniqlash uchun ishlab chiqilgan qoidalar hamda taklif qilingan usullar sinovdan o'tkazilganida omonim so'z orasidan keng qo'llaniladiganlari saralab olinib ularning har biri bo'yicha o'rtacha



10-rasm. Axborot tizimining konseptual modeli



11-rasm. Axborot tizimining umumiy biznes-jarayonlarining BPM modeli

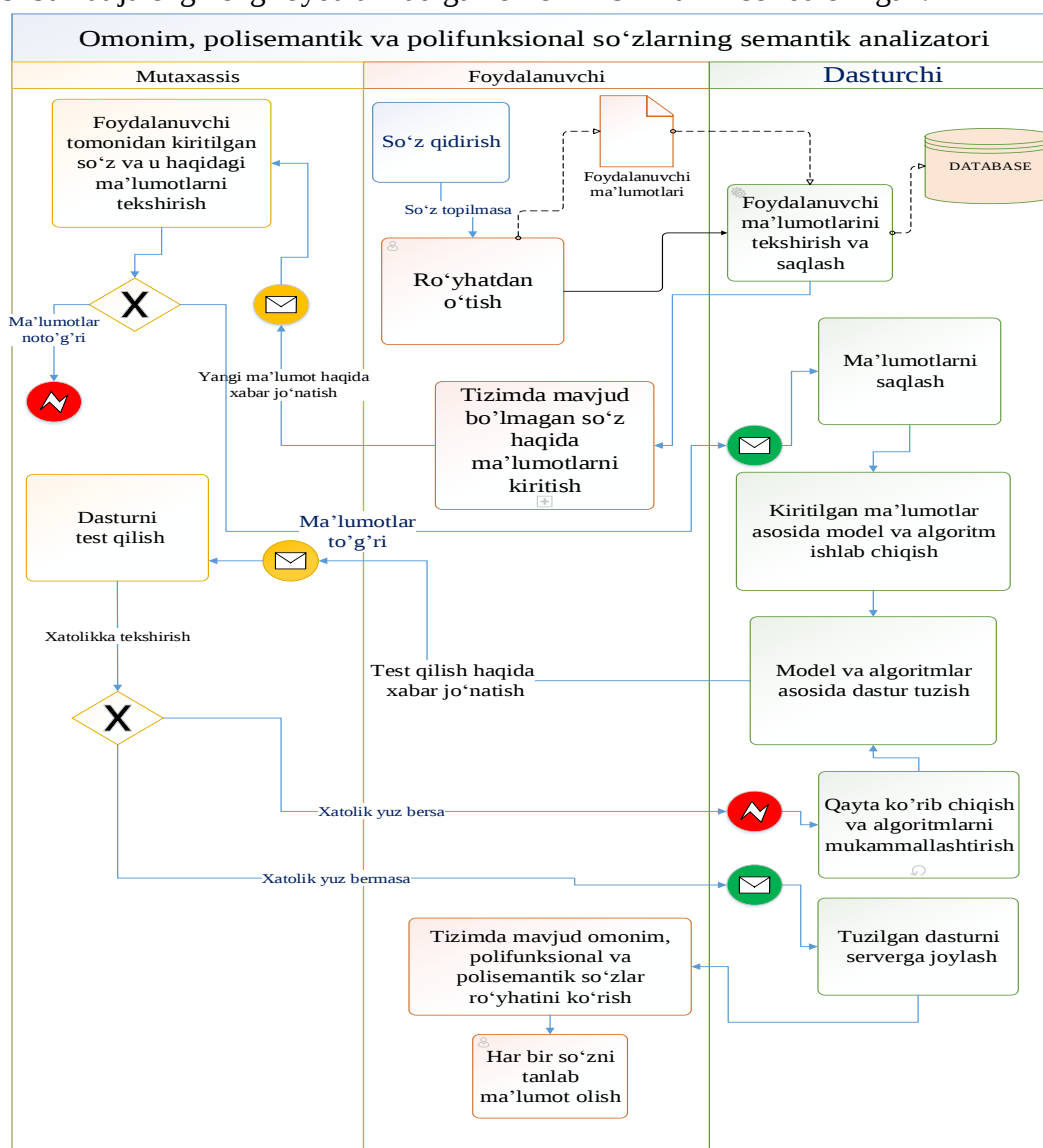
**Omonim, polisemantik va polifunksional soʻzlarda foydalanilgan
usullardan olingan natijalar**

Usullar	Omonimiya		Poli- semiya	Polifunk- sionallik (120)
	Bir soʻz turkumi doirasidagi omonimiya (692)	Turli soʻz turkumlari doirasidagi omonimiya (406)		
		Grammatik jihatdan oʻxshash boʻlgan soʻz turkumlari orasidagi omonimiya (243)	Grammatik jihatdan oʻxshash boʻlmagan soʻz turkumlari orasidagi omonimiya (163)	
Qoidalarga asoslangan usul	–	–	81,7%	–
Chastotali usul Naïve Bayes usuli	–	95%	7,3%	12,1%
YMM	80,5%	93%	–	70%

192 ta gap tizimga kiritildi va ularda 157 holatda aniq natija olindi.

$$\varphi = \frac{n}{m} = \frac{157}{192} = 0,81,7$$

Ushbu natija eng keng foydalaniladigan omonim soʻzlar misolida olingan.



12-rasm. Tizimda joriy etilgan rollar mexanizmining BPM modeli

XULOSA

O'zbek tilidagi omonim, polifunksional va polisemantik so'zlarni semantik analiz qiluvchi axborot tizimini ishlab chiqishda qator qoidalar ishlab chiqildi, axborot tizimidagi har bir funksional jarayon belgilab olindi ularning BPM modellari ishlab chiqildi. O'zbek tilidagi polisemantik so'zlarni kontekstda farqlash uchun albatta so'zlar orasidagi bog'lanishlar muhim ahamiyat kasb etadi. Bu bog'lanishlar turkiy tillarda Grammatik aloqa deb ataladi. Gap tarkibidagi so'zlarning o'zaro aloqasi yordamida semantik analizatorlar ishlab chiqiladi.

Mazkur tadqiqot ishida quyidagi natijalarga erishildi:

1. O'zbek tilidagi omonim so'zlarning grammatik xususiyatlaridan kelib chiqib ikkita guruhga bir *so'z turkumi doirasidagi omonimlar* hamda *turli so'z turkumidagi omonimlarga* ajratib farqlash muhim ahamiyat kasb etishi aniqlandi. Bir so'z turkumidagi omonimlikni hosil qiluvchi so'zlar 5 (faqat ot, fe'l, sifat, olmosh, ravish) guruhga ajratilib, ular Mashinali o'qitishga asoslangan yondoshuvning Yashirin Markov modeli yordamida aniqlanishi asoslandi va natijalar olindi.

2. Turli so'z turkumlari doirasidagi omonimlikni hosil qiluvchi omonimlarni qaysi so'z turkumlari doirasida uchrashiga ko'ra 37 ta guruhga ajratilgan va ularni farqlashda qoidalarga asoslangan usul va Chastotali usul, hamda Naive Bayes usullaridan foydalanilgan. Turli so'z turkumlari doirasidagi omonim so'zlarni grammatik jihatdan farqli bo'lganlari ajratib olindi, ularni farqlovchi matematik modellar tavsiya qilindi. Xuddi shunday polisemantik va polifunksional so'zlarni grammatik xususiyatlariga ko'ra semantik farqlovchi matematik modellar tavsiya qilindi. Matematik modellarni ishlab chiqishda 72 ta belgilashlardan foydalanildi. Foydalanilgan belgilar jahon tajribasiga asoslangan.

3. O'zbek tilidagi omonim, polifunksional va polisemantik so'zlarni semantik farqlovchi axborot tizimining vazifasiga ko'ra uning konseptual modeli, axborot modeli hamda ma'lumotlar bazasining normallashtirish qoidalariga asoslangan ma'lumotlar bazasining ER modeli taklif etilgan. ER modelida 23 ta jadval tasvirlangan.

4. Axborot tizimidagi omonim, polisemantik hamda polifunksional so'zlarni semantik farqlovchi biznes-jarayonlarning BPM modellari ishlab chiqilgan. Omonim so'zlarni farqlovchi 6 ta, polifunksionallikni farqlovchi 4 ta, polisemiyaning farqlovchi 1 ta, axborot tizimini xususiyatlarini ifodalovchi 3 BPM modellari ishlab chiqildi. Axborot tizimining funksional xususiyatlaridan kelib chiqib, uning arxitekturasini hamda strukturasi ishlab chiqilgan.

5. Ishlab chiqilgan matematik modellar, modellashtirilgan funksional jarayonlar asosida o'zbek tilidagi gaplar tarkibidagi omonim, polisemantik hamda polifunksional so'zlarni semantik analiz qiluvchi axborot tizimi ishlab chiqildi hamda sinovdan o'tkazildi. O'zbek tilidagi eng ko'p foydalaniladigan omonim, polisemantik hamda polifunksional so'zlar saralab olinib, axborot tizimi sinovdan o'tkazilganda 81.7% aniqlikka erishildi. Axborot tizimining aniqlik darajasini oshirish uchun neyron tarmoq modellaridan foydalanish tavsiya etildi.

**НАУЧНЫЙ СОВЕТ DSc.13/30.12.2019.Т.07.01 ПО ПРИСУЖДЕНИЮ
УЧЕНЫХ СТЕПЕНЕЙ ПРИ ТАШКЕНТСКОМ УНИВЕРСИТЕТЕ
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ**

**ТАШКЕНТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
УЗБЕКСКОГО ЯЗЫКА И ЛИТЕРАТУРЫ ИМЕНИ АЛИШЕРА НАВОИ**

АХМЕДОВА ХОЛИСХОН ИЛХОМОВНА

**РАЗРАБОТКА МОДЕЛИ, АЛГОРИТМОВ И
ИНФОРМАЦИОННОЙ СИСТЕМЫ СЕМАНТИЧЕСКОГО АНАЛИЗА
ПРЕДЛОЖЕНИЙ УЗБЕКСКОГО ЯЗЫКА**

05.01.10 – Информационные системы и процессы

**АВТОРЕФЕРАТ ДИССЕРТАЦИИ
ДОКТОРА ФИЛОСОФИИ (PhD) ПО ТЕХНИЧЕСКИМ НАУКАМ**

The topic of the Doctor of Philosophy (PhD) dissertation in technical sciences is registered with the Higher Attestation Commission under the Ministry of Higher Education, Science and Innovation of the Republic of Uzbekistan under the number B2023.1. PhD/T2289.

Диссертация выполнена в Ташкентский государственный университет узбекского языка и литературы имени Алишера Навои.

Аннотация диссертации на трех языках (узбекский, русский, английский (резюме)) размещен на веб-странице (www.tuit.uz) и на Информационно-образовательном портале «Ziynet» (www.ziynet.uz).

Научный руководитель:

Элов Ботир Болтаевич

доктор философии по техническим наукам, доцент

Официальные оппоненты:

Сандов Абдусобиржон Абдурахмонович

доктор технических наук, профессор

Мухамедиева Дильноз Тулкуновна

доктор технических наук, профессор

Ведущая организация:

**Самаркандский государственный университет
имени Шарофа Рашидова**

Защита диссертации состоится «22» декабря 2023 г. в 14:00 часов на заседании научного совета DSc.13/30.12.2019.T.07.01 при Ташкентском университете информационных технологий. (Адрес: 100084, г. Ташкент, ул. Амира Темура, 108. Тел.: (99871) 238-64-43; e-mail: iktuit@tuit.uz).

С диссертацией можно ознакомиться в Информационно-ресурсном центре Ташкентского университета информационных технологий (регистрационный номер № 2857). (Адрес: 100084, г. Ташкент, ул. Амира Темура, 108. Тел.: (99871) 238-64-70).

Аннотация диссертации разослан «7» декабря 2023 года.
(протокол рассылки № 22 от «3» ноября 2023 г.).



M. M. Musaev

М.М. Мусаев
Председатель Научного совета
по присуждению учёных степеней,
доктор технических наук, профессор

H. O. Rakhimov

Н.О. Рахимов
Ученый секретарь Научного совета
по присуждению учёных степеней,
доктор технических наук, доцент

D. B. Sultanov

Дж.Б. Султанов
Председатель научного семинара при Научном
совете по присуждению ученых степеней,
доктор технических наук, доцент

ВВЕДЕНИЕ (аннотация диссертации доктора философии (PhD))

Актуальность и востребованность темы диссертации. В результате усилий по автоматизации существующих задач, упрощения и облегчения человеко-машинного интерфейса за счет широкого внедрения современных информационных технологий в сферу компьютерной лингвистики уже в XXI веке люди изучают естественные языки и анализируют предложения посредством информационных систем. Такие средства являются продуктом новейших информационных и коммуникационных технологий, включая системы технической обработки (processing) и распознавания (recognition) изображений и речи, системы морфологического, синтаксического анализа текстов. Безусловно, в области компьютерной лингвистики большое внимание уделяется результатам исследований, проводимых ведущими учеными таких стран как США, Франция, Германия, Великобритания, Китай, Южная Корея, Япония, Россия, которые являются передовыми с точки зрения современных технологий.

В мире существует множество анализаторов, которые выполняют семантический анализ предложений на различных естественных языках. Все используемые при этом методы и средства различаются в зависимости от лексических и грамматических особенностей выбранного языка. В этом плане, важными задачами являются ведение научных изысканий по предварительной обработке текстов и разработке лингвистических анализаторов, разработка ER (Entity Relationship) модели базы данных информационных систем, осуществляющих целостное и безупречное с лингвистической точки зрения выражение особенностей естественного языка, выбор оптимальных подходов и решений для моделирования бизнес-процессов обработки языка и разработки информационных систем.

В Республике на государственном уровне осуществляется широкое использование современных информационных технологий и их исследование, популяризация разрабатываемых информационных систем, поддержка исследований в данной области. В частности, в Указе Президента Республики Узбекистан «Об утверждении стратегии «Цифровой Узбекистан – 2030» и мерах по ее эффективной реализации», в том числе определены такие задачи как «...проведение фундаментальных и прикладных исследований по приоритетным направлениям в области информационных технологии вперед... внедрение перспективных инновационных разработок и start-up проектов и поддержка их коммерциализации и трансферта технологий... развитие и стимулирование исследований и разработок в области цифровых технологий, совершенствование их организационных механизмов...»³. В целях реализации этих задач, в том числе анализа сложных систем, производимых на основе самых современных на сегодня технологий и разработки узбекской модели человеко-машинного интерфейса при их использовании, информационных

³ Указ Президента Республики Узбекистан от 5 октября 2020 года №УП-6079 «Об утверждении Стратегии «Цифровой Узбекистан-2030» и мерах по ее эффективной реализации».

систем, в частности предложений на узбекском языке, на основе исследований, проводимых в рамках различных проектов. Разработка и представление информационных систем семантического анализа является одной из важных задач сегодняшнего дня.

Данное диссертационное исследование в определенной степени служит реализации задач, предусмотренных Указом Президента Республики Узбекистан от 30 июня 2017 года №УП-5099 «О мерах по коренному улучшению условий для развития отрасли информационных технологий в республике», Постановлениями Президента Республики Узбекистан от 20 апреля 2017 года №ПП-2909 «О мерах по дальнейшему развитию системы высшего образования», от 6 октября 2020 года №ПП-4851 «О мерах по дальнейшему совершенствованию системы образования в области информационных технологий, развитию и интеграции научных исследований с IT-индустрией», Указом Президента Республики Узбекистан от 20 октября 2020 года №УП-6084 «О мерах по дальнейшему развитию узбекского языка и совершенствованию языковой политики в стране», а также другими нормативно-правовыми документами, относящимися к данной сфере деятельности.

Соответствие исследования приоритетным направлениям развития науки и технологий республики. Данное исследование выполнено в соответствии с приоритетным направлением развития науки и технологий IV – «Информатизация и развитие инфокоммуникационных технологий».

Степень изученности проблемы. Исследования многих ученых мира посвящены ряду проблем в сфере NLP (Natural Language Processing). В мировой компьютерной лингвистике учеными проведены научные изыскания по вопросам семантического анализа и анализатора. К их числу можно отнести работы таких ученых как А.В.Тузов, М.В.Мозговой, А.В.Сокирко, N.A.Schlayefer, А.В.Мочалова, А.Р.Гатиатуллин, В.Yergesh, A.Sharipbay, Г.Бекманова, S.Lipniski, Y.A.Kanevskiy, К.К.Боярский.

Теоретические и практические работы по автоматизации узбекского языка нашли отражение в работах таких ученых как М.Мусаев, М.Шарипов, Х.Мадатов, Ш.Бекчанов, Б.Элов, И.Бакаев, Б.Акмурадов, У.Хамдамов, Ж. Элов, У.Салаев. Делаются первые шаги по разработке семантического анализатора на основе лингвистических моделей, предложенных учеными-языковедами.

Связь темы диссертационного исследования с научно-исследовательскими работами высшего учебного заведения, в котором выполняется диссертация. Диссертационное исследование выполнено в рамках инновационного проекта IL-402104209 «Разработка программного средства морфологического анализатора и анализа морфолексикона узбекского языка – средства автоматической обработки для информационно-поисковых систем (Google, Yandex, Google translate)» (2022-2024) Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои.

Целью исследования разработка информационной системы, базы данных и структур, моделирование процессов семантического анализа полисемантических и полифункциональных слов, омонимов для семантического анализатора узбекского языка.

Задачи исследования:

изучение и анализ научно-исследовательских работ, процессов информатизации в сфере обработки естественного языка и особенностей внедренных информационных систем;

разработка концептуальной модели информационной системы семантического анализа предложений на узбекском языке;

разработка ER-модели базы данных и информационной модели информационной системы, семантически дифференцирующей омонимы, полисемантические и полифункциональные слова узбекского языка на основе правил нормирования базы данных;

определение методов и разработка правил, семантически дифференцирующих омонимы, полисемантические и полифункциональные слова, исходя из их особенностей;

разработка моделей бизнес-процессов BPM (Business Process Model), дифференцирующих омонимы, полисемантические, полифункциональные слова для информационной системы, семантически различающей предложения на узбекском языке;

разработка архитектуры, структуры информационной системы семантического анализатора и моделирование функциональных процессов;

разработка информационной системы семантического анализа предложений на узбекском языке.

Объектом исследования являются процессы и системы семантического анализа омонимов, полифункциональных и многозначных слов в узбекском языке.

Предмет исследования – методы, модели и процессы дифференциации омонимов, полифункциональных и полисемантических слов.

Методы исследования. В ходе исследования использованы метод проектирования информационных систем и реляционных баз данных, метод структурного и функционального моделирования, метод разработки информационных систем на основе моделирования бизнес-процессов.

Научная новизна исследования заключается в следующем:

определены методы дифференциации в контексте омонимичных, полифункциональных и многозначных слов в узбекском языке и разработана концептуальная модель информационной системы семантического анализатора;

на основе реляционной модели данных и правил нормализации сбора и обработки данных с использованием основных атрибутов базы данных построена ER-модель базы данных информационной системы, семантически разграничивающей омонимичные, полифункциональные и многозначные слова узбекского языка;

разработаны ВРМ-модели бизнес-процессов, семантически различающих омонимичные, многозначные и полифункциональные слова в узбекском языке, процессы обеспечены лингвистическими данными;

на основе статистических данных разработана информационная система определения омонимии, многозначности и полифункциональности в предложениях, сбора, обработки и представления информации, способствующей принятию решений.

Практические результаты исследования заключаются в следующем:

разработаны концептуальная модель информационной системы семантического анализатора, ВРМ-модели функциональных процессов;

разработана ER-модель базы данных информационной системы, позволяющей дифференцировать омонимы, полифункциональные и полисемантические слова узбекского языка;

разработаны информационные модели информационной системы, позволяющей дифференцировать омонимию, полисемию и полифункциональность;

разработан интерфейс, удобный для дифференцирования в корпусе омонимии, полифункциональных и многозначных слов, ввода, изменения, удаления и представления данных пользователям.

Достоверность результатов исследования заключается в том, что изученные материалы способствуют получению выводов, исходя из природы узбекского языка, их обоснованность опирается на лингвистические основы узбекского языка, источники, подтвержденные на практике при разработке процессов, моделей семантического анализатора узбекского языка.

Научная и практическая значимость результатов исследования. Исследование обладает теоретической и практической значимостью при ведении исследований по компьютерной лингвистике при разработке теоретических основ разработки семантического анализатора и его математических моделей. Позволяет усовершенствовать национальный корпус узбекского языка и повысить его уровень.

Внедрение результатов исследования. На основе значимости семантического анализатора в информационно-поисковой системе, математических моделей дифференциации омонимов, полифункциональных и многозначных слов, математических моделей, бизнес-процессов и информационных систем, осуществляющих дифференциацию омонимов, полифункциональных и многозначных слов для семантического анализатора узбекского языка:

в результате внедрения и испытания комплекса алгоритмов и математических моделей, разработанных в ходе исследования, созданы всего 19 математических моделей, дифференцирующих в общей сложности 1065 омонимов, 1 модель на основе математической логики, различающая многозначные слова, и 5 математических моделей, семантически дифференцирующая полифункциональные слова;

для семантической дифференциации омонимов и многозначных слов в узбекском языке разработаны более 200 групп слов, более 50 групп аффиксов и комбинаций аффиксов;

обоснованы способы использования статистических методов на основе корпусных данных для определения омонимии;

разработана ERD-модель базы данных информационной системы, осуществляющей определение омонимии, полисемии и полифункциональности;

в ходе исследования на основе “Метода на основе правил”, “Метода на основе статистических данных” и “Метода на основе машинного обучения”, примененных в разработке информационной системы, осуществляющей семантический анализ предложений на узбекском языке, достигнуто решение вопроса омонимии (свидетельство Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои от 22 ноября 2022 года). В результате внедрения и испытания разработанного комплекса бизнес-процессов и математических моделей выявлены методы, разработаны 19 математических моделей, которые позволяют дифференцировать около 3000 омонимичных слов, существующих в узбекском языке, 1 модель на основе математической логики, различающая многозначные слова, и 5 математических моделей, семантически дифференцирующих полифункциональные слова; на основе «Метода на основе правил», «Частотного метода», «Наивного байесовского классификатора» и скрытых марковских моделей, использованных в информационной системе, которая разработана в ходе исследования, и разработанных бизнес-процессов точность определения устранения омонимии составила 84%, точность различения полисемии 73% и точность дифференциации полифункциональности 74,5% соответственно;

результаты исследования использованы при разработке программы «MyDoctor Consultant», созданной программистами софт-фирмы ООО «Mustafo Software», размещенной в городе Ургенче Хорезмской области (свидетельство директора ООО «Mustafo Software» от 1 февраля 2023 года). По итогам использования результатов исследования достигнуто определение многозначности слов, обозначающих части тела человека. При определении значений в представленных выводах использованы результаты выявления семантических сходств. Это позволило повысить в 1,2 раза эффективность программы «MyDoctor Consultant».

Апробация результатов исследования. Результаты исследования обсуждены на 8 международных и 5 Республиканских конференциях.

Опубликованность результатов исследования. По теме исследования опубликовано всего 30 научных работ, из них 8 статей в научных изданиях, рекомендованных Высшей аттестационной комиссией к публикации основных научных результатов диссертаций, 4 статьи в зарубежных журналах (из них 2 в базе данных Scopus), издана 1 монография.

Структура и объем диссертации. Диссертация состоит из введения, трех

глав, заключения, списка использованной литературы и приложений. Объем диссертации составляет 118 страниц.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обоснованы актуальность и востребованность темы диссертации, определены цель и задачи, объект и предмет исследования, приводится соответствие исследования приоритетным направлениям развития науки и технологий Республики Узбекистан, изложены научная новизна, практические результаты исследования, обоснована достоверность полученных результатов, раскрывается теоретическая и практическая значимость результатов исследования, приведены внедрение результатов исследования, сведения об опубликованности результатов и структуре диссертации.

Первая глава диссертации получила название **«Методы и системы анализа предложений»**. В первом параграфе главы *«Системы анализа предложений в естественных языках»* проанализированы общие вопросы анализатора предложения, его виды, исследования анализатора мировой и узбекской компьютерной лингвистики. В параграфе также описаны общие сведения о таких видах лингвистического анализатора как *морфологический, синтаксический, семантический анализаторы*, а также их функции.

Во втором параграфе главы *«Модели и методы семантического анализа предложений»* описаны модели и методы семантического анализа текстов. Системы, осуществляющие семантический анализ на разных языках, основаны на разных моделях и алгоритмах.

В главе предложены математические выводы для разработки системы семантического анализа омонимов, полисемантических и полифункциональных слов. Существующие подходы к определению омонимии основаны на локальном и глобальном синтаксическом анализе, синтаксических словарях и вероятностях. Явление омонимии определяется на основе свойств слов в корпус с использованием грамматических и статистических явлений. Для того, чтобы омонимия была успешной нужны точные результаты. Кроме того, большой объем данных замедляет их обработку, это называется **комбинаторным взрывом**. В семантическом анализе предложений в мировой компьютерной лингвистике используются 3 метода: *Подход на основе правил; Подход на основе статистических данных; Подход на основе машинного обучения*.

Подход на основе правил – предполагает семантический анализ слов с помощью правил грамматики языка. В узбекском языке разработан ряд правил. На основе этих правил разработаны математические умозаключения, которые позволяют дифференцировать соответствующие виды омонимов. Сущность данного метода заключается в том, что в некоторых случаях анализ контекста помогает понять синтаксическую структуру части предложения, что позволяет понять формы слова.

Подход на основе статистических данных – используется посредством классификации грамматических параметров слова при решении задач. Эти

параметры выбираются разными способами в разных естественных языках. К параметрам классификации можно такие параметры как морфологический анализ, лемма, корень, корень + аффикс. При невысоком уровне точности семантического анализа возникает необходимость в расширении параметров классификации. Основной задачей данного метода является выделение n-граммов контекста, т.е. оцениваются сочетания введенного слова, которые оцениваются методами оценки. В некоторых методах статистики решения принимаются посредством слова и его аффиксов, в других же вывод делается на основе семантической валентности (сочетаемости) в контексте слова.

Подход на основе машинного обучения – основан на использовании нейронных сетей. Данный метод является дополнением к статистическим методам, сейчас широко применяется учеными мира при определении омонимии. С помощью метода на основе машинного обучения можно определить омонимию в рамках одной части речи в узбекском языке. Зарубежный опыт показал, что для использования модели нейросети необходимы соответствующие тексты, предварительно определенные. ИМз этого следует, что актуальной проблемой является определение значений, тегирование текстов большого объема в базе данных национального корпуса узбекского языка.

В третьем параграфе главы *«Методы семантической разметки текстов в семантическом анализаторе узбекского языка»* классифицированы вопросы тегирования текстов, используемых при создании семантического анализатора узбекского языка. Последовательное моделирование проблем процесса обработки естественных языков (NLP-Natural languages processing) является требованием времени. Определение в тексте омонимов и полифункциональных слов сталкивается непосредственно с проблемой тегирования частей речи. Например, в процессе определения значений омонимичных слов в рамках двух частей речи в узбекском языке в соответствующем контексте значимыми являются *Метод тегирования на основе правил*, *Метод стохастического тегирования*.

Тегирование на основе правил. При автоматическом тегировании частей речи обычно подходы на основе правил используют контекстные сведения для записи тегов к неизвестным словам.

Метод стохастического тегирования. Термин СТ «Stochastic tagging» предполагает применение разных подходов к проблеме тегирования. В некоторых источниках применяется также статистический метод. Каждая модель, включающая в себя частотность или вероятность в любом случае, называется *корректной стохастической моделью*. Самое простое стохастическое тегирование выделяет слова на основе вероятности появления слова с определенным тегом.

Вторая глава диссертации получила название **«Методы семантического анализа слов»**. В первом параграфе главы *«Методы устранения омонимии в узбекском языке»* изучены математические модели семантической дифференциации омонимов в узбекском языке. Омонимичные слова в

узбекском языке семантически можно разделить на две группы: *омонимы в рамках одной части речи* и *омонимы в рамках разных частей речи*. Разработаны математические выводы по дифференциации омонимы в каждой группе. Классификация омонимов в рамках частей речи приводится в рис. 1. Отмечается, что в исследованиях Ш.Гулямовой, посвященных наблюдению омонимии в рамках каждой части речи, т.е. глагола, существительного, прилагательного, междометия и местоимения, что нашло отражение в рис. 2.

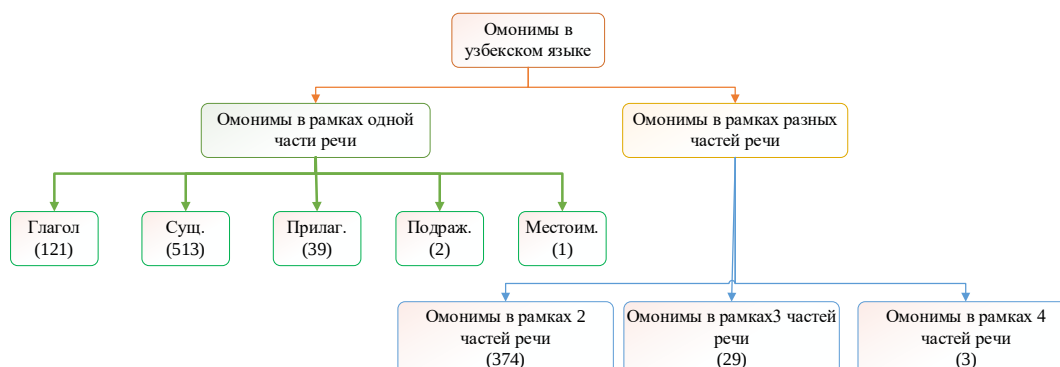


Рис. 1. Иерархия омонимов в узбекском языке

Определение омонимии в рамках каждой части речи можно привести на примере глагола *suzmoq*, которое имеет 5 значений в рамках части речи глагол. Слово **suz** в разном контексте имеет разные значения, при этом наблюдается множество сочетаний с одинаковыми аффиксами. При решении данной задачи рекомендуется применять метод на основе машинного обучения, поскольку не существует определенного правила при дифференциации омонимии в рамках одной части речи. Если же есть, но не имеет общий характер. Также обосновано невозможность решения данной задачи с помощью статистических методов. Следовательно, определение омонимии в рамках одной части речи предполагает обращение к *нейронным сетям на основе машинного обучения*. В ходе исследования диссертант не обращался к нейронным сетям. Но предложено при решении проблемы применять другой метод определения омонимии в рамках одной части речи, т.е. употребить группы слов, которые необходимы при использовании нейронных сетей. Очевидно, что с помощью этого метода нельзя полностью достичь поставленной цели, но можно решить задачу на определенном уровне. Определение омонима посредством семантической валентности в мире получило название метод *семантического конкорданса*. Определены x_i значения глагола *suzmoq* (рис. 2).

X_1	≡	Suyuqlikni filtrlamoq
X_2	≡	Suvda harakatlanmoq
X_3	≡	Ovqatni idishga solmoq
X_4	≡	Noz-karashma qilmoq

Рис. 2. Значения слова **suz**

X_1	⇒	Suyuqlik, hayvon, uy ajnom, inson
X_2	⇒	Hayvon, inson, suyuqliklar
X_3	⇒	Ovqat, oshxona anjom, idish-tovoq
X_4	⇒	Inson ismlari, ko'z, hayvon, shaxsga ishora qiluvchi so'zlar. nigoh bilan bog'liq fe'llar

Рис. 3. Группы слов семантических сочетаний слова **suz**

Следовательно, существуют 5 скрытых случаев, их неявность заключается в том, встречающийся в заданном предложении глагол *suzmoq*, неизвестно какое из значений использовано. Для определения значения глагола *suzmoq* в заданном предложении используется контекст. Для этого следует предварительно определить валентность (сочетаемость) омонима. При сочетании омонима *suz* с семантическими сочетаниями определены значения (см.: рис. 4).

В данном рисунке каждое слово обладает собственной семантической иерархией. Следовательно, можно прийти к выводу о том, что для определения значения слова с помощью контекста необходимо рассматривать слова узбекского языка, выделяя их в семантические группы. Сочетания глагола *suzmoq* на рис. 4 *сузмок*, относящиеся к таким группа слов как *sut emizuvchi hayvonlar*, *filtrlanish xususiyatiga ega bo'lgan suyuqliklar*, *suyuqlik bilan bog'liq bo'lgan harakat va holat fe'llari*, *oshxona anjomlari*, имеют значение *suyuqlikni filtrlamoq* (фильтровать жидкость).

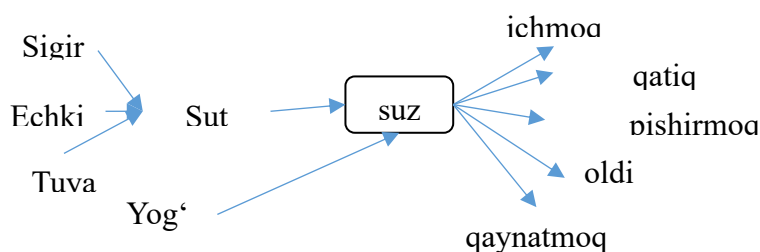


Рис. 4. Совокупность семантической валентности глагола *suzmoq*

Здесь может возникнуть вопрос: *Слов, которые могут быть омонимами в рамках одной части речи много, следует ли для каждого из них определять такого рода группы слов?* Для поиска ответа на этот вопрос диссертантом проведены изыскания, проанализировано множество исследований, изучен мировой опыт. Автор пришел к выводу, что нет другого решения данной проблемы. Хотя возникла мысль о необходимости выделения омонимов в семантические группы, но это себя не оправдало. Для определения семантической валентности каждого омонима, относящегося к одной части речи, и выделения их в семантические группы необходимы произведения, статьи, сведения большого объема. Такого рода сведения есть в базе данных Национального корпуса узбекского языка⁴, в котором размещены миллиарды текстов. Специалистами отмечается, что заполнение групп слов словами, дифференциация и определение сочетаний омонимичных слов с помощью корпусных данных. В параграфе главы «Методы устранения омонимии в рамках разных частей речи» освещены вопросы использования методов на основе правил и на основе статистических данных.

Выделены три группы в соответствии с образованием омонимии в рамках разных частей речи в узбекском языке (в рамках 2, 3, 4 частей речи) и определены группы частей речи. При дифференциации омонимии в рамках разных частей речи использован метод на основе правил. В качестве примера

⁴ <http://uzschoolcorpara.uz/uz>

приводится модель дифференциации слов, составляющих омонимию в рамках трех частей речи (существительное v числительное v глагол):

1. **Uch:** Biror narsaning boshlanish yoki tugash qismi (часть начала или конца чего-либо). (*существительное*)
2. **Uch:** Ikki bilan to'rt orasidagi butun son (Целое число между два и четыре). (*числительное*)
3. **Uch:**
 1. Havo bo'ylab harakatlanmoq (двигаться по воздуху). (*глагол*)
 2. Biror qirrasi shart uzilib tushmoq (обрыв края чего-либо).
 3. Beixtiyor berilmoq, laqillamoq (невольно отдаваться).

В ходе наблюдений собраны следующие правила:

- **uch**+*-i/-iga/-ni/-da* $\oplus$ *bor/ulamoq/boshlamoq/turmoq/...* (*существительное*)
- **uch**+*-inchi/-nchi/-ta/-tacha/-tadan/-larcha/* $\oplus$ *oy nomlari/dars/uy/...* (*числительное*)
- **uch**+*-ov/-ovlon/...* $\oplus$ *gaplashmoq/ko'rishmoq/kelmoq/ketmoq/...* (*числительное*)
- **uch**+*dan* $\oplus$ *bir/ikki/uch/to'rt/...* (*числительное*)
- *Balandga/osmonga/...* $\oplus$ **uch** (*глагол*)
- *Toshkentdan/shahardan/...* $\oplus$ **uch** (*февл*)

Учитывая приведенные морфологические особенности, можно привести следующую модель дифференциации омонимов в части речи числительное.

$$H_{num,N,V} = \begin{cases} H^{num} + \downarrow aff_l^{ord} \oplus \downarrow (W_k^N + \downarrow aff_l^{ord}) \oplus N_j^c \\ \downarrow W_i^{dec} \oplus H^{num} + \downarrow aff_l^{ord} \oplus N_j^c \\ H^{num} + aff_j^{agg} \oplus V \\ H^{num} + dan \oplus W_n^{num} \\ N + aff_i^N \\ V + aff_j^V \end{cases} \quad (1)$$

При этом aff_j^{agg} – множество общих аффиксов, $aff_j^{agg} = -ov, -ovlon, -ala, -ovi, -oviga, -ovloni, \dots$; W_n^{num} – множество слов, обозначающих названия чисел, $W_n^{num} = \{bir, ikki, uch, \dots\}$

Среди разных частей речи имеются части речи, которые обладают сходствами в грамматическом аспекте (рис. 5). Применение строгих правил при определении омонимии среди них не дает результата. Поэтому целесообразно использовать метод на основе статистических данных. Частотный метод является одним из таких статистических методов. Рассмотрено определение омонимии в рамках таких частей речи как *существительное* или *прилагательное* с помощью частотного метода. Для использования омонимии требуется выделить параметры классификации, позволяющие выделить части речи.



Рис. 5. Части речи, которые имеют грамматические сходства и образуют омонимию.

Омонимы в рамках частей речи *существительное* V *прилагательное* классифицируются по следующим параметрам:

1. Часть речи;
2. Корень и лемма;
3. Только лемма;
4. Только корень

При поиске через сайт uzschoolcorpora.uz слово *alomat*, образующее омонимию в рамках частей речи *прилагательное* V *существительное* всего найдено 6823 результатов, анализ 100 из них показал, что 79 относятся к *существительному*, а 21 – к *прилагательному* соответственно. Анализ найденных результатов с дифференциаций корня и аффиксов позволил осуществить расчет следующих вероятностей.

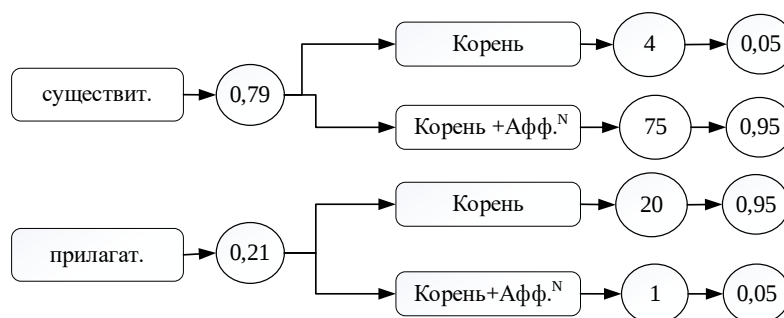


Рис. 7. Вероятности появления слова *alomat* в корпусе

Анализ вероятностей, рассчитанных на основе параметров классификации, показал, что слова, которые образуют омонимию в рамках частей речи *существительное* или *прилагательное*, оцениваются как относящиеся к *существительным*, если в составе слова в 90% случаев имеются сочетания, которые свойственны *существительным*. Кроме того, затрагиваются вопросы использования метода Байеса при определении омонимии в рамках разных частей речи. Для слова *alomat*, приведенного в качестве примера, образуются значения двух условных вероятностей.

В третьем параграфе главы «Методы выявления многозначных слов в узбекском языке» рассматриваются математические модели дифференциации полисемии. В данном разделе работы также разработана математическая модель определения полисемии на примере слова *ko'z*. Обосновано применение метода сегментации при определении полисемии.

Anor xolaning ko'zi tekkan aslida, aks holda o'qishga kirgan bo'lardi.

В приведенном примере многозначное слово *ko'z* употреблено в сочетании со словом *xola*, но использование слова *tekkan* после *ko'z* ведет к переносу значения. Из этого следует, что в зависимости от сочетаний с другими словами многозначное слово, образующими прямое и переносное значения, следует выделять соответствующие группы. Необходимо выделять группы слова на основе представленной иерархии сочетаний, которые образуют прямое и переносное значения слова *ko'z*. Чем оптимальнее будет распределение слов по малым группам, тем точнее будет семантического анализа многозначных слов. В соответствии с зарубежным опытом при семантическом анализе многозначных слов лучшим решением будет использование методов статистики.

В зависимости от прямого и переносного значений сочетаний со словом **ko'z** выделяются группы.

W_i^b – P множество слов, передающих прямое значение в сочетании с предыдущим от полисемантического слова словом;

W_j^a – P множество слов, передающих переносное значение в сочетании с последующим от полисемантического слова словом.

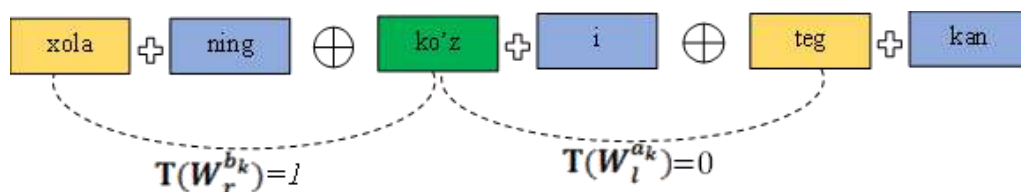
Представляя множества выделенных сочетаний как существующие, приводится следующее математическое выражение

$$W_r^{b_k} + \downarrow aff^{s-l} \oplus P_k + \downarrow aff^{s-l} \oplus W_l^{a_k} + \downarrow aff^{s-l} = \begin{cases} P_k^s, T(W_r^{b_k}) \wedge T(W_l^{a_k}) = 0 \\ P_k^p, T(W_r^{b_k}) \wedge T(W_l^{a_k}) = 1 \end{cases} \quad (2)$$

Рассмотрен анализ слова *ko'z* в приведенной ниже фразе с помощью модели (3).

Anor xolaning ko'zi tekkan aslida, aks holda o'qishga kirgan bo'lardi.

Слово *ko'z* в данной фразе сочетается с последующим словом *xola* (обладающая способности видеть). Следовательно, относится к группе $xola \in \{inson\}$ и $T(xola)=1-xola$ сочетается со словом *ko'z*, что передает прямое значение (орган зрения) многозначного слова; предшествующее слово *teg* не относится к функции *ko'z* и поэтому $T(teg)=0$. Диссертант пришел к выводу,



что для того, чтобы достичь высокого результата с помощью математической модели (3) для слова *ko'z* в приведенном предложении необходимо выделить существующие в узбекском языке слова в семантические группы.

$$T(xola) \wedge T(teg) = 0$$

$$1 \wedge 0 = 0$$

В третьем параграфе главы «*Модели семантической дифференциации полифункциональных слов в контексте*» приводится классификация полифункциональных слов в узбекском языке. Полифункциональные слова так же как и омонимы могут быть дифференцированы на основе морфологических свойств. Но применение этого метода недостаточно для выявления полифункциональных слов, составления модели. Поскольку некоторые части речи определяются только из контекста. Тем не менее, необходимы ясные правила, модели для системы определения полифункциональности. Ш.Гулямова классифицируя слова в разрезе частей речи, разработала лингвистическую модель для полифункциональных слов каждой группы. На основе данных лингвистических моделей разработаны лингвистические модели для каждой группы полифункциональных слов. На основе этих лингвистических моделей уже были разработаны математические модели. Приводится математическая модель, определяющая полифункциональность среди частей речи *существительное* $\vee$ *прилагательное*.

– Kasal+ -i/ni/ga/lar/da/ini/... $\oplus$ aniqlandi/ko 'rmoq/keldi/...

– Kasal $\oplus$ odam/bemor/hayvon/qush/...

$$Pf_{adj,N} = \begin{cases} Pf^N + \downarrow aff^N \oplus V \\ Pf^{adj} + \downarrow aff^{deg} \oplus N \end{cases} \quad (3)$$

Рекомендовано определять полифункциональность среди частей речи *существительное* $\vee$ *прилагательное* с помощью (3). Точно также разработаны математические модели, определяющие полифункциональность среди таких частей речи как *прилагательное* $\vee$ *модальные слова*, *прилагательное* $\vee$ *наречие*, *самостоятельный глагол* $\vee$ *вспомогательный глагол*, *союз* $\vee$ *модальные слова*.

$$Pf_{M,Adj} = \begin{cases} Pf^{adj}, \text{ если } Pf^{M,Adj} + \downarrow aff^{adj} \oplus W^{Pf_{adj}} \\ Pf^M, \text{ если } W^M \oplus Pf^{M,Adj} \oplus W^M \end{cases} \quad (4)$$

$$Pf_{Adj,Adv} = \begin{cases} Pf^{Adj}, \text{ если } Pf^{Adj,Adv} + aff^{deg} \oplus N^{Pf_{adj}} \\ Pf^{Adv}, \text{ если } Pf^{Adj,Adv} \oplus V \end{cases} \quad (5)$$

$$Pf_{I,Aux} = \begin{cases} Pf^I, \text{ если } W^x + aff^{CS} \oplus Pf^{I,L} \\ Pf^{Aux}, \text{ если } V^{Aux} + aff^{Aux} \oplus Pf^{I,Aux} \end{cases} \quad (6)$$

$$Pf_{Conj,M} = \begin{cases} Pf^{Conj}, \text{ если } W^{Conj} \oplus Pf^{Conj,Aux} \oplus W^{Conj} \\ Pf^M, \text{ если } W^M \oplus Pf^{Conj,M} \oplus V \end{cases} \quad (7)$$

Третья глава диссертации «**Модели и информационной системы и проектирование процессов семантического анализатора узбекского языка**» состоит из трех параграфов. В первом разделе главы, озаглавленном «*Проектирование информационной модели информационной системы семантического анализа предложения в узбекском языке*» информационная модель, лежащая в основе информационной системы (рис. 7) и модель ER

Разработана база данных, рассмотрены ее объекты, информационные типы данных. База данных ER-модель информационной системы семантического анализа предложений на узбекском языке была разработана с использованием нотации IDEF1X на платформе Visio (рис. 8).

Во втором параграфе главы «*Моделирование функциональных процессов информационной системы семантического анализа предложений в узбекском языке*» разработаны концептуальная модель информационной системы семантического анализатора, BPM-модели бизнес-процессы информационной системы, осуществляющей семантический анализ омонимов, полисемантических и полифункциональных слов в предложениях на узбекском языке. В рис. 9 приводится концептуальная модель информационной системы семантического анализатора. При определении омонимии учитывается также вопросы образования омонимии посредством имен собственных и лексических единиц, обозначающих человека. Разработаны алгоритмы для каждой модели, дифференцирующие омонимию в рамках разных частей речи.

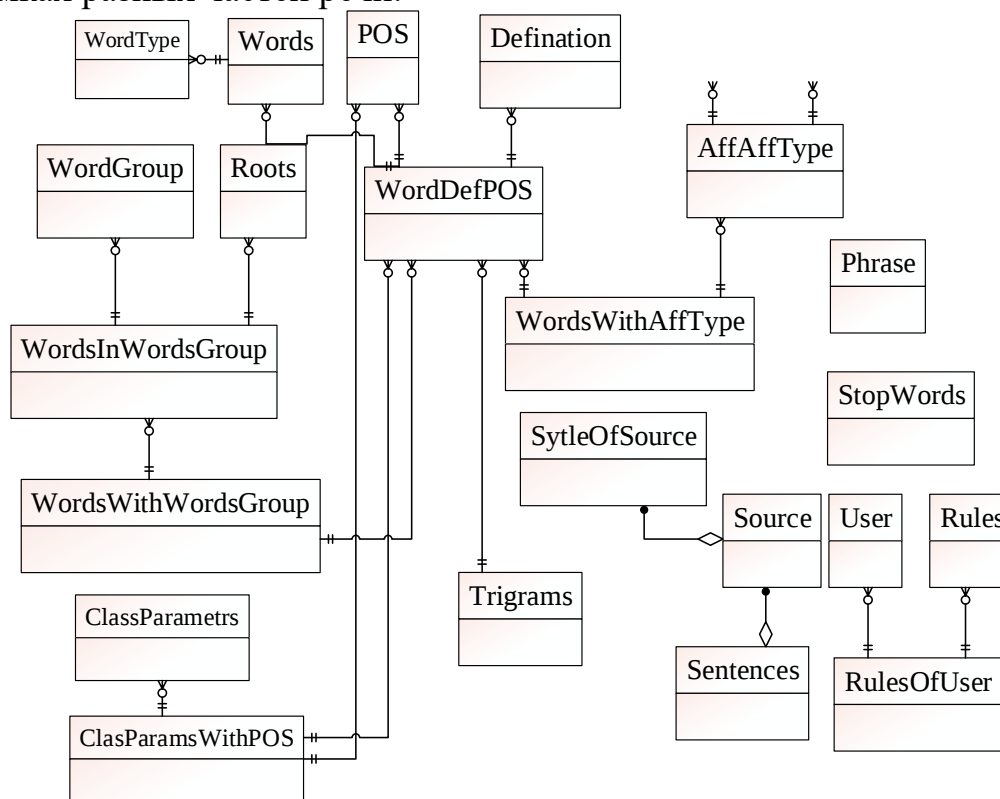


Рис. 7. Информационная модель информационной системы семантического анализа предложений на узбекском языке

В третьем параграфе главы «*Возможности и аналитические результаты информационной системы семантического анализатора*» рассмотрены возможности информационной системы, определяющей значения омонимов, полифункциональных и полисемантических слов в составе предложений в узбекском языке. Кроме того, приводятся данные о семантическом анализе и его изучении. Семантический анализ имеет большое значение для NLP, поскольку его процессы позволяют определить разные значения слов. Вместе

с тем, эти процессы способствуют пониманию значений целых фраз и текстов. При разработке информационной системы использованы возможности фреймворка Django языка программирования Python. Пользователи информационной системы семантического анализатора обращаются к сети Интернет и локальной сети с помощью веб-браузера. В этом плане, для обеспечения безопасности необходимо обеспечить аутентификацию пользователей, аутентификацию веб-сервера, обеспечить конфиденциальность трафика от веб-браузера до веб-сервера. Архитектура аутентификации, идентификации и авторизации пользователей системы отражена в рис. 10. Используются механизм ролей в информационной системе для обеспечения авторизации пользователей. В рис. 12 определены функции ролей.

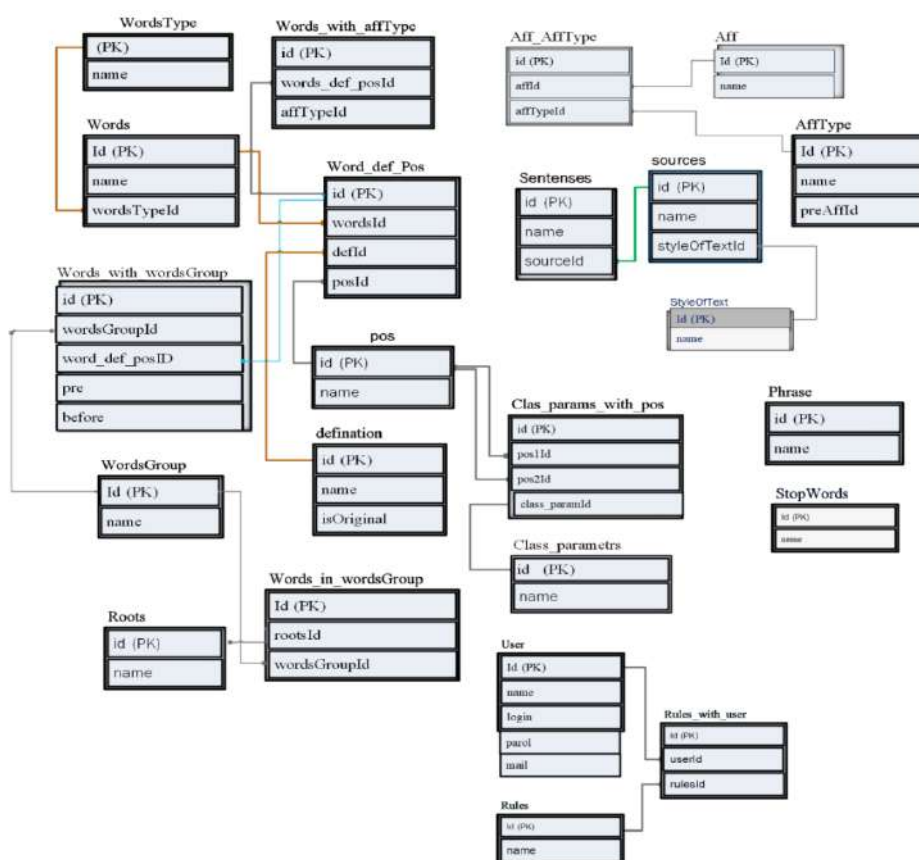


Рис. 8. ER-модель базы данных информационной системы семантического анализатора

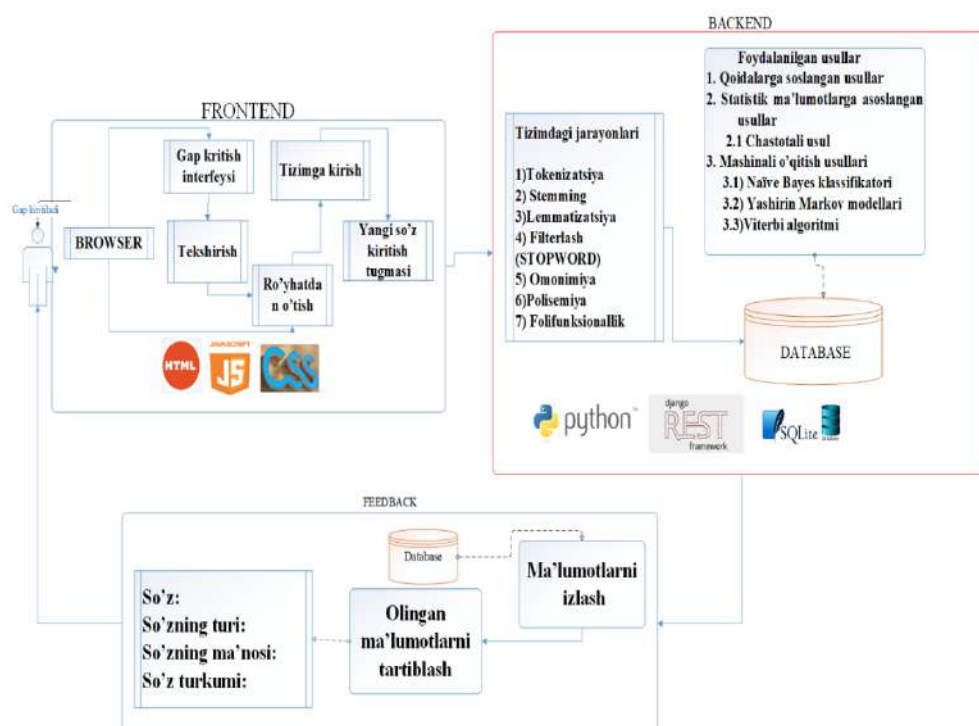


Рис. 9. Концептуальная модель информационной системы



Рис. 10. Аутентификация, идентификация и авторизация пользователей семантического анализатора

Уровень точности методов, использованных при разработке информационной системы семантического анализатора, основывался на том же, что и в таблице 1. Приведенные в таблице 1 значения получены по результатам применения математических моделей при определении слова каждого типа, на основе выявленных параметров классификации и семантических групп слов. При разработке информационной системы семантического анализа предложений на узбекском языке использованы Метод на основе правил, Частотный метод, Наивный Байесовский классификатор и Скрытый марковский метод. Для использования скрытого марковского метода в базу данных размещены около 5000 семантически тегированных словосочетаний, состоящих из триграмм существующих омонимов. При испытании предложенных методов и разработанных правил для определения омонимии отобранны наиболее часто употребляемые омонимы, которые в среднем в 192 системы предложений и в 157 случаях получены точные результаты.

$$\varphi = \frac{n}{m} = \frac{157}{192} = 081,7$$

Данный результат получен на примере широко используемых омонимов, если анализ проводится по всем существующим омонимам, то показатели будут ниже

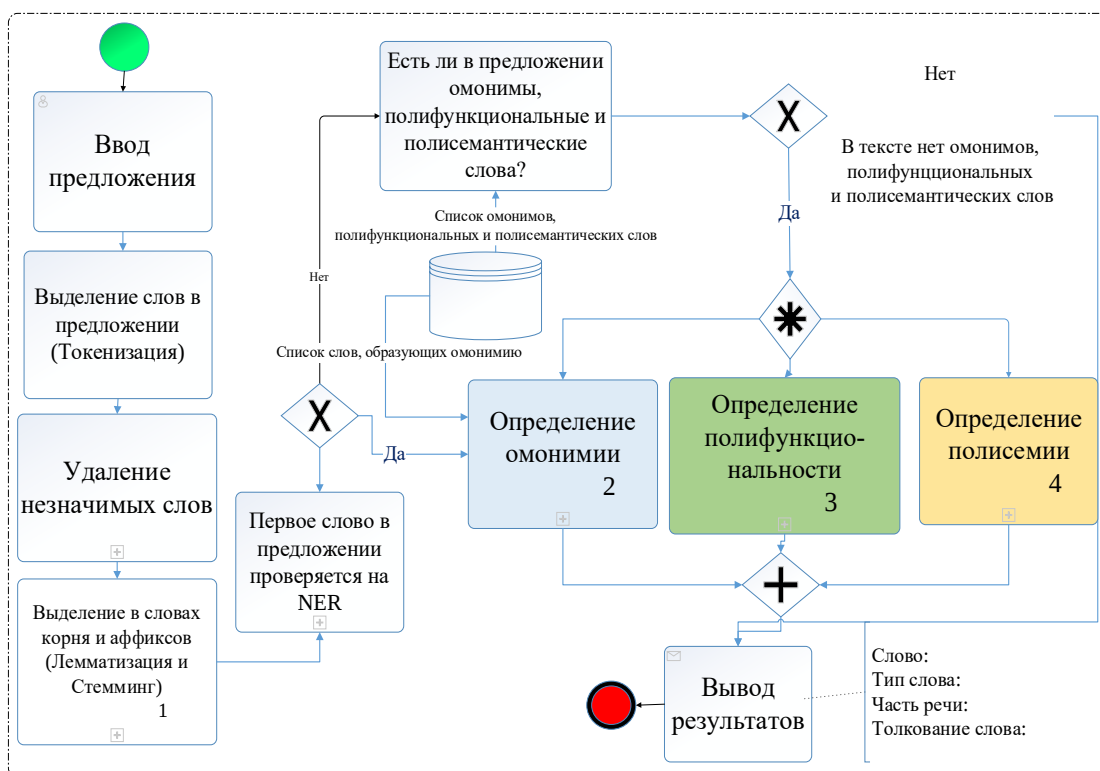


Рис. 11. Общая BPMN-модель бизнес-процессов информационной системе

Таблица 1.
Результаты, полученные в ходе использования методов в омонимах, полисемантических и полифункциональных словах

Методы	Омонимия			Полисемия	Полифункциональность (120)
	Омонимия в рамках одной части речи (692)	Омонимия в рамках разных частей речи (406)			
		Омонимия среди грамматически схожих частей речи (243)	Омонимия среди грамматически несхожих частей речи (163)		
Метод на основе правил	—	-	81,7%	-	71,7%
Частотный метод и Наивный байесовский классификатор	—	95%	7,3%	12,1%	-

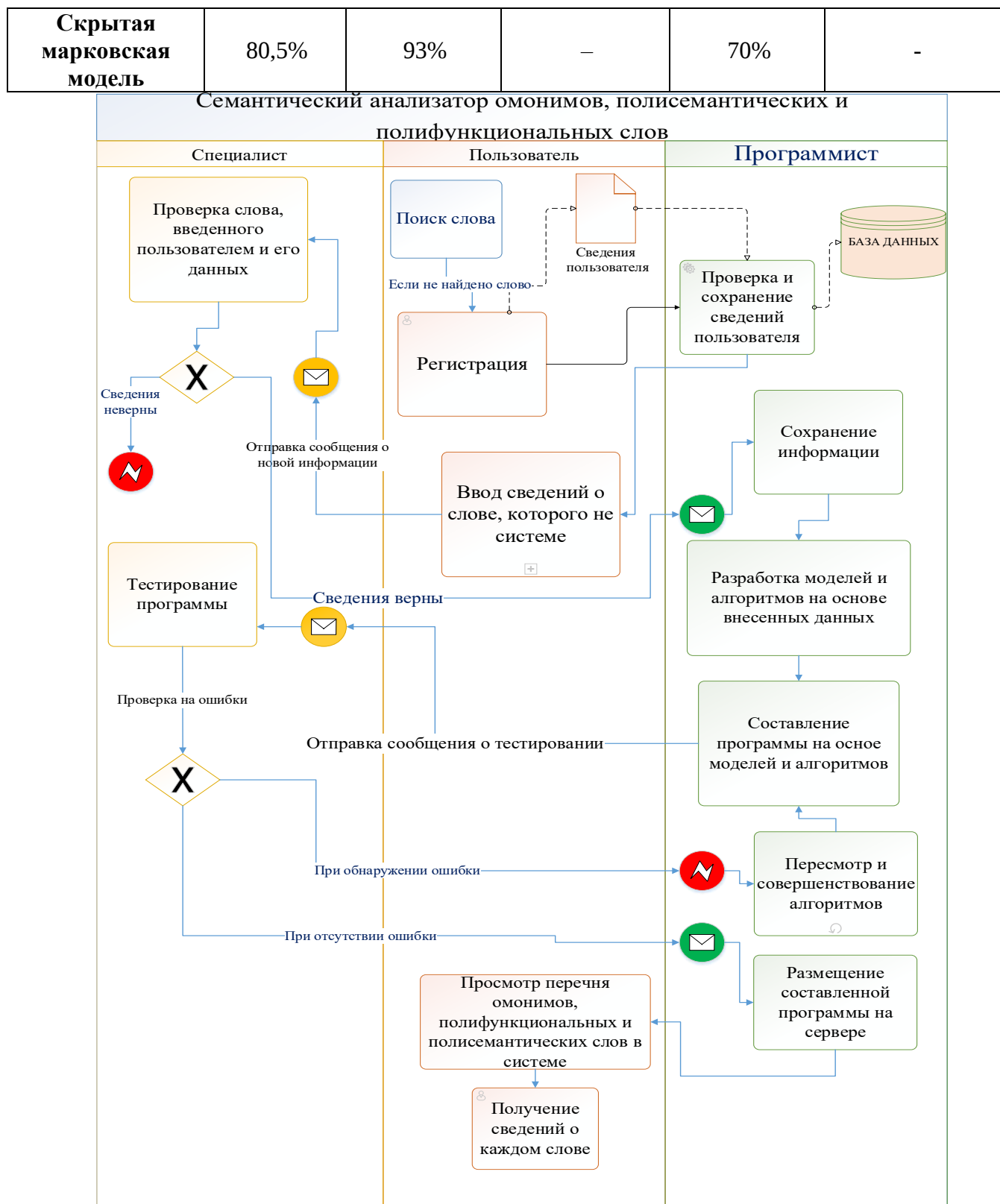


Рис. 12. BPM-модель механизма роллера, внедренного в систему

ЗАКЛЮЧЕНИЕ

Разработаны правила создания информационной системы по семантическому анализу омонимов, полифункциональных и полисемантических слов узбекского языка, определен каждый функциональный процесс в информационной системе, разработаны их BPM-

модели. Для определения полисемантических слов узбекского языка в контексте имеют значение связи между словами. Данные связи в тюркских языках называются Грамматической связью. Разработаны семантические анализаторы на основе взаимосвязи слов в предложении.

В ходе исследования получены следующие результаты:

1. Выявлено, что, исходя из грамматических свойств омонимов в узбекском языке, важное значение имеет выделение их в две группы: *омонимы в рамках одной части речи* и *омонимы в рамках разных частей речи*. Слова, образующие омонимию в рамках одной части речи, выделены в 5 групп (только существительное, прилагательное, глагол, местоимение, наречие), обосновано их определение и получение результатов с помощью скрытой марковской модели подхода на основе машинного обучения.

2. Слова, образующие омонимию в разных частях речи, выделены в 37 групп в зависимости от того, к каким частям речи они относятся. При их дифференциации использованы Метод на основе правил, Частотный метод и Наивный байесовский классификатор. Омонимы в рамках разных частей речи выделяются в группу грамматически разных слов. Рекомендованы математические модели их дифференциации. Аналогично были рекомендованы математические модели, семантически различающие многозначные и полифункциональные слова по их грамматическим признакам. При разработке математических моделей использовано 72 обозначения. Используемые символы основаны на мировом опыте.

3. В соответствии с функцией информационной системы, семантически дифференцирующей омонимичные, полифункциональные и многозначные слова в узбекском языке, предложены ее концептуальная модель, информационная модель и ER-модель базы данных на основе правил нормирования базы данных. В ER-модели представлены 23 таблицы.

4. Разработаны BPM-модели бизнес-процессов, семантически различающие в информационной системе омонимичные, многозначные и полифункциональные слова. Разработано 6 BPM-моделей для дифференциации омонимов, 4 для различения полифункциональности, 1 для разграничения полисемии и 3 для представления особенностей информационной системы. На основе функциональных характеристик информационной системы разработаны ее архитектура и структура.

5. На основе разработанных математических моделей и смоделированных функциональных процессов разработана и испытана информационная система семантического анализа омонимичных, многозначных и полифункциональных слов в предложениях на узбекском языке. Отсортированы и протестированы наиболее употребляемые омонимичные, многозначные и полифункциональные слова в узбекском языке, достигнута точность 81,7% при испытании информационной системы. Для повышения уровня точности информационной системы рекомендовано использовать нейросетевые модели.

**SCIENTIFIC COUNCIL AWARDING SCIENTIFIC DEGREES
DSc.13/30.12.2019.T.07.01 AT TASHKENT UNIVERSITY OF
INFORMATION TECHNOLOGIES**

**TASHKENT STATE UNIVERSITY OF UZBEK LANGUAGE AND
LITERATURE NAMED AFTER ALISHER NAVOI**

AKHMEDOVA XOLISKHON ILKHOMOVNA

**DEVELOPMENT OF MODEL ALGORITHMS AND INFORMATION
SYSTEM FOR SEMANTIC ANALYSIS OF SENTENCES IN THE UZBEK
LANGUAGE**

05.01.10 – Information systems and processes

**DISSERTATION ABSTRACT OF THE DOCTOR OF PHILOSOPHY
(PhD)
ON TECHNICAL SCIENCE**

Tashkent-2023

The topic of the Doctor of Philosophy (PhD) dissertation in technical sciences is registered with the Higher Attestation Commission under the Ministry of Higher Education, Science and Innovation of the Republic of Uzbekistan under the number B2023.1. PhD/T2289.

The dissertation has been prepared at Alishera Navoi, Tashkent State University of Uzbek Language and Literature.

Scientific adviser: **Elov Botir Boltaeovich**
Doctor of Philosophy (PhD) on Technical Sciences, docent

Official opponents: **Saidov Abdusobirjon Abdurakhmonovich**
Doctor of Technical Sciences, Professor

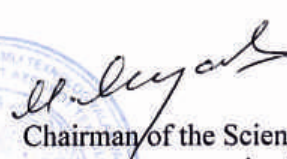
Mukhamedieva Dilnoz Tulkunovna
Doctor of Technical Sciences, Professor

Leading organization: **Samarkand State University named after Sharof Rashidov**

Defense of the dissertation will take place on « 22 » december 2023 at _____ at a meeting of the Scientific Council DSc.13/30.12.2019.T.07.01 at the Tashkent University of Information Technologies (Address: 100084, Tashkent city, Amir Temur street, 108. Ph: (+99871) 238-64-43, e-mail: iktuit@tuit.uz).

The doctoral dissertation could be reviewed in Information-Resource Center of Tashkent University of Information Technologies (registration number № 2857). (Address: 100084, Tashkent city, Amir Temur street, 108. Ph: (99871) 238-64-70,

The abstract of dissertation is distributed on « 7 » december 2023 y.
(Dispatching protocol № 22 on « 3 » november 2023y.)



M.M. Musayev
Chairman of the Scientific Council awarding scientific degrees, Doctor of Technical Sciences, Professor


N.O. Rakhimov
Scientific Secretary of Scientific Council awarding scientific degrees, Doctor of Technical Sciences, Docent


Dj.B. Sultanov
Chairman of the Scientific Seminar at the Scientific Council awarding scientific degrees, Doctor of Technical Sciences, Docent

INTRODUCTION (abstract of PhD dissertation)

The aim of the research work is to develop an information system, data and structures for modeling the processes of semantic analysis of polysemantic and polyfunctional words, homonyms for the semantic analyzer of the Uzbek language.

The object of research is homonyms, polyfunctional and polysemantic words of the Uzbek language.

The scientific novelty of the research work:

methods for distinguishing homonymous, polyfunctional and polysemantic words in the Uzbek language in context were determined and a conceptual model of the semantic analyzer information system was developed;

based on the relational data model and the rules for normalizing data collection and processing using the main attributes of the database, an ER-model of the information system database was built, semantically delimiting homonymous, polyfunctional and polysemantic words of the Uzbek language. developed;

an ER model of an information system database has been developed, BPM models of business processes have been developed that semantically distinguish between homonymous, polysemantic and polyfunctional words in the Uzbek language, and the processes are provided with linguistic information;

Based on statistical data, an information system has been developed for identifying homonymy, polysemy and polyfunctionality in sentences, collecting, processing and presenting information to help make decisions.

Implementation of the research results. The significance of the semantic analyzer in the information retrieval system, mathematical models for the differentiation of homonyms, polyfunctional and polysemantic words, mathematical models, business processes and information systems that differentiate homonyms, polyfunctional and polysemantic words for the semantic analyzer of the Uzbek language are used in:

a software package that performs the semantic analysis of homonyms, polyfunctional and polysemantic words as part of sentences in the Uzbek language, was introduced into the activities of a state institution. The Center for Training on the Basics of Office Work in the State Language and Advanced Training at the Tashkent State University of the Uzbek Language and Literature named after Alisher Navoi (certificate of the Tashkent State University of the Uzbek language and literature named after Alisher Navoi of November 22, 2022). As a result of the implementation and testing of a set of developed mathematical models and business processes, mathematical models for differentiating about 3,000 homonyms that exist in the Uzbek language, 1 model based on mathematical logic that differentiates polysemantic words and 5 mathematical models for the semantic differentiation of polyfunctional words have been developed; the accuracy of determining homonymy based on the “Method based on rules”, “Method based on statistical data”, “Method of semantic concordance”, applied in the software package that was developed during the study, and business processes is 88%, the accuracy of differentiating polysemy and polyfunctionality make up 82% and 86% respectively;

the results of the study were used in the development of the program "MyDoctor Consultant", created by programmers of the software company Mustafo Software LLC, located in the city of Urgench, Khorezm region (certificate of the director of Mustafo Software LLC dated February 1, 2023). As a result of using the results of the study, a definition of ambiguity among the parts of the human body was achieved. When determining the values in the presented conclusions, the results of identifying semantic similarities were used. This made it possible to increase the effectiveness of the MyDoctor Consultant program by 1.2 times.

Structure and size of the dissertation. The dissertation consists of an introduction, three chapters, a conclusion, a list of references and applications. The volume of the dissertation is 118 pages.

E'LON QILINGAN ISHLAR RO'YXATI

I bo'lim (I часть; I part)

1. Elov B.B., Axmedova X.I. O'zbek tili semantik analizatorining matematik modellari.- Monografiya. – Toshkent: Farovon yurt nashriyoti, 2021.
2. Elov B.B., Axmedova X.I. Business Process Modeling That Distinguishes Homonymy Within Three Parts of Speeches in Uzbek Language// Computer Engineering Department Heads (BMBB) meetings, 7th edition of the conference, UBMK'22, Dicle University, Diyarbakir,Turkiya on September 14-15-16, 2022. (№3, Scopus)
3. Elov B.B., Hamroyeva Sh.M., Axmedova X.I. Methods for creating a morphological analyzer// The 14th International Conference on Intelligent Human-Computer Interaction (IHCI-2022), October 20th-22nd, 2022 Tashkent, Uzbekistan (№3, Scopus; №11, Springer)
4. Axmedova X. I. Turli so'z turkumlari orasidagi omonimiyani aniqlovchi matematik modellar// Science and innovation international scientific journal volume 1 issue 7 uif-2022: 8.2 | issn: 2181-3337 Tashkent, Uzbekistan. <https://doi.org/10.5281/zenodo.7238546> (№43,UIF; №23,SJIF)
5. Axmedova X.I. Mathematical models that distinguish homonymy in the framework of a word series // Electronic journal of actual problems of modern science, education and training october, 2021-10/1. ISSN 2181-9750,46-51-b (Urganch) (05.00.00 №26)
6. Elov B. B., Axmedova X.I. HOMONYMY DETECTION USING A NAÏVE BAYES CLASSIFIER Journal of Computer Science Engineering and Information Technology Research (JCSEITR) ISSN(P): 2250-2416; ISSN(E): Applied Vol. 13, Issue 1, March 2023, 5–20 © TJPRC Pvt. Ltd, Hindiston (05.00.00 №28)
7. 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 innovasion rivojlanish ilmiy jurnal 2022 / 1, 150-162-b. Toshkent. (05.00.00, OAK rayosatining 2019-y.28-fevraldagi 262/9.2-sonli qarori)
8. Elov B.B., Axmedova X.I. Business process modeling that distinguishes homonymy within three part of speeches //NeuroQuantology, May 2022 , Volume 20, Issue 5, Page 2176-2188, eISSN 1303-5150. doi: 10.14704/nq.2022.20.5.NQ22578 (№3, Scopus)

II bo'lim (II часть; II part)

9. Elov B.B., Axmedova X.I. A Mathematical Model That Semantically Analyzes Polysemantic Words// Journal of Pedagogical Inventions and Practices in Volume 3. <http://www.zienjournals.com/>, 119-122-b.(№23, SJIF; №14, ReasearchBib)

10. Axmedova X.I., Masharipov S.A. Distinguish homonyms using statistical methods// “O‘zbekistonda ilmiy-amaliy tadqiqotlar ” Ilm sarchashmalari jurnali, Urganch, 11. 2021 y., 149-153-b.
11. Elov B.B., Axmedova X.I. Polisemantik so‘zlarni semantik tahlil qilishning matematik modellari // Til va madaniyat 2021 vol.1 (3),Toshkent,77-86-b.
12. Elov B.B., Axmedova X.I. Determining homonymy using statistical methods// “Hisoblash modellari va texnologiyalari (HMT 2022)” O‘zbekiston-Malayziya ikkinchi xalqaro konferensiyasi materiallari- Toshkent, 2022 16-17 sentabr,-106 b.
13. Elov B.B., Axmedova X.I. Omonim so‘zlarni farqlash usullari // “O‘zbek milliy va ta’limiy korpuslarini yaratishning nazariy hamda amaliy masalalari” Xalqaro ilmiy-amaliy konferensiya, Vol. 1 №. 01 (2021) Toshkent, 41-44-b.
14. Gulyamova Sh.Q., Axmedova X.I. Omonim so‘zlarning lingvistik asosi, matematik modeli va algoritmlari (Ot va fe'l hamda sifat va fe'l turkumi doirasida). – Qo‘qon DPI. Ilmiy xabarlar. – Qo‘qon, 2021. – № 2. – B. 103-110
15. Gulyamova Sh.Q., Axmedova X.I. O‘zbek tili semantik analizatori uchun omonim so‘zlarlar ma'lumotlar bazasini shakllantirish masalasi xususida// So‘z san'ati xalqaro jurnali. 4-jild, 3-son. – Toshkent, 2021.– B. 326-334. – Doi 10.26739/2181-9297
16. Axmedova X.I. O‘zbek tilidagi shakldosh so‘zlarni semantik tahlil qilish // “O‘zbekistonning umidli yoshlari” mavzusidagi 7-son Respublika ilmiy talabalar, magistrlar yosh tadqiqotchilar va mustaqil izlanuvchilar uchun onlayn konferensiyasining materiallari to‘plami.,O‘zbekiston 2021. 39-41-b.
17. Axmedova X.I. Ravish/fe‘l va ravish/yuklama so‘z turkumlari orasidagi omonimlikni farqlovchi matematik modellar // “O‘zbek tili taraqqiyoti va xalqaro hamkorlik masalalari” mavzusidagi xalqaro konferensiya (2021-yil 18-oktabr), Toshkent, 455-457-b.
18. Axmedova X.I., Xusainova Z.Y. Fe‘l/ko‘makchi va sifat/son so‘z turkumlari orasidagi omonimlikni farqlovchi matematik modellar// “O‘zbek tili taraqqiyoti va xalqaro hamkorlik masalalari” mavzusidagi xalqaro konferensiya (2021-yil 18-oktabr), Toshkent, 458-460-b.
19. Axmedova X.I. Bir so‘z turkumi doirasidagi omonimlikni farqlovchi matematik modellar// “O‘zbek amaliy tilshunosligi va lingvodidaktikasi masalalari” mavzusidagi xalqaro ilmiy-amaliy konferensiya materiallari (2021-yil 20-oktabr), Toshkent 153-158-b.
20. Axmedova X.I. Ot/taqlid so‘z turkumlari orasidagi omonimlikni farqlovchi matematik modellar// “O‘zbek amaliy tilshunosligi va lingvodidaktikasi masalalari” mavzusidagi xalqaro ilmiy-amaliy konferensiya materiallari (2021-yil 20-oktabr), Toshkent 177-179-b.
21. Axmedova X.I. Polisemantik so‘zlarni kontekstda semantik tahlil qilish modellari// “O‘zbekistonda ilmiy-amaliy tadqiqotlar ” mavzusidagi respublika 34-ko‘p tarmoqli ilmiy masofaviy onlayn konferensiya materiallari to‘plami 16-qism, 2021-yil, Toshkent,27-30-b.

22. Elov B.B., Axmedova X.I. O‘zbek tilidagi polifunksional so‘zlarni semantik farqlovchi biznes-jarayonlarni modellashtirish// Til va madaniyat 2022 vol.1 (5), Toshkent, 87-100-b.

23. Axmedova X.I. Omonim so‘zlarni farqlovchi biznes-jarayonlarni modellashtirish: ikki so‘z turkumi doirasidagi omonim so‘zlar misolida// “Kompyuter lingvistikasi: muammolar, yechim, istiqbollar” xalqaro ilmiy-amaliy konferensiya vol. 1 №. 01 (2022). Toshkent. 479-484-b.

24. Axmedova X.I. Korpus ma’lumotlari yordamida bir so‘z turkumi doirasidagi omonimiyani aniqlovchi parametrlarni hisoblash //“Kompyuter lingvistikasi: muammolar, yechim, istiqbollar” xalqaro ilmiy-amaliy konferensiya vol. 1 №. 01 (2022). Toshkent, 485-492-b.

25. Axmedova X.I. Chastotali usul yordamida omonimiyani aniqlash// “O‘ZBEK AMALIY FILOLOGIYASI ISTIQBOLLARI” Respublika ilmiy-amaliy konferensiyasi Toshkent: 2022.-164-170 b.

26. Elov B.B., Axmedova X.I. Turli so‘z turkumlari orasidagi omonimiyani na’ive bayes kalssifiktori yordamida aniqlash//“O‘zbek tilining milliy korpusi: muammolar va vazifalar” xalqaro ilmiy-amaliy Anjuman, 2023 yil 25 mart, Samarqand, 251-257-b.

27. Axmedova X.I. Statistik usullar yordamida turli so‘z turkumlari orasidagi omonimiyani aniqlash// Uzbekistan: Til va Madaniyat 2023/1 ISSN 2181-922X/ 91-105-b.

28. Axmedova X.I. Importance of statistical methods in determine the homonymy// «Компьютер лингвистикасиннинг замонавий технологиялари - CTCL.2023» (<https://www.myscience.uz/index.php/issue/issue/view/12>) mavzusida vazirlik miqyosidagi ilmiy-amaliy konferentsiya, Toshkent 14.04.2023, 210-215 p.

29. Mengliev B., Elov B., Gulyamova Sh., Axmedova X. Ot va fe’l, sifat va fe’l so‘z turkumlari orasidagi omonimlikni farqlovchi lingvistik, matematik modellar va vositalarning ma’lumotlar bazasi // – № DGU 12071 11.08.2021.

30. Gulyamova Sh., Axmedova X. Bir so‘z turkumi doirasidagi omonimlikni farqlovchi lingvistik, matematik modellar va vositalarning ma'lumotlar bazasi // № DGU 20213100

Avtoreferat «Muhammad al-Xorazmiy avlodlari» ilmiy jurnali tahririyatida
tahrirdan o'tqazildi va o'zbek, rus, ingliz tillaridagi matnlarini mosligi
tekshirildi

