ISNN 3030-3907

2026/4
VOLUME 1

DEVELOPMENT OF SCIENCE

ILMIY JURNAL



Development of science

Ilmiy jurnal

2026/4 VOLUME 1

ISSN 3030 -3907

Ilmiy jurnal OAK rayosatining 2024 yil 27 sentabrdagi 361-son qaroriga asosan 02.00.00 kimyo fanlari hamda 2025 yil 12 fevral 367- son qaroriga asosan 05.00.00 texnika fanlari bo‘yicha falsafa doktori (PhD) va fan doktori (DSc) ilmiy darajasiga talabgorlarning dissertatsiya ishlari yuzasidan dissertatsiyalari asosiy ilmiy natijalarini chop etish tavsiya etilgan milliy ilmiy nashrlar ro‘yxatiga kiritilgan.

BOSH MUHARRIR:
Temirov Alisher Hoshim o‘g‘li – texnika fanlari falsafa doktori (PhD),Dotsent.
TAHRIRIYAT KENGASH RAISI:
Fozilov Sadriddin Fayzullayevich – texnika fanlari doktori, professor
TAHRIRIYAT KENGASH A’ZOLARI

TEXNIKA FANLARI/ TECHNICAL SCIENCES/ ТЕХНИЧЕСКИЕ НАУКИ

Bozorov G‘ayrat Rashidovich
Buxoro Davlat Texnika Universiteti, professor
Abdurahmonov Olim Rustamovich
Buxoro Davlat Texnika Universiteti, professor
Adizov Bobur Zamirovich
O‘zbekiston Respublikasi Fanlar akademiyasi Umumiy va noorganik kimyo instituti, professor
Hayitov Ruslan Rustamjonovich
Buxoro Davlat Texnika Universiteti, professor
Jumayev Qayum Karimovich
Buxoro Davlat Texnika Universiteti, professor.
Berdiyev Dorob Murodovich
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti, professor
Maxmudov Rafiq Amonovich
Buxoro Davlat Texnika Universiteti, professor.
Tilloyev Lochin Ismatilloyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Raxmonov Qahramon Sanakulovich
Buxoro Davlat Texnika Universiteti, dotsent.
Barakayev Nusratullo Rajabovich
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti, professor.
Jo‘rayev Xayrullo Fayziyevich
Buxoro Davlat Texnika Universiteti, professor.
Panoyev Erali Rajabboyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Ataullayev Sherzod Nabiullyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Murodov Malik Negmurodovich
Buxoro Davlat Texnika Universiteti, dotsent.
Axmedova Ozoda Bahronovna
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti.
Safarov Bahri Jumayevich
Buxoro Davlat Texnika Universiteti, dotsent.
Xaydarov Axtam Amonovich
Buxoro Davlat Texnika Universiteti, dotsent.
Safarov Jasur Alijon o‘g‘li
Buxoro Davlat Texnika Universiteti, dotsent.
Асадова Раъно Дилмуратовна
Тошкент Давлат техника университети доценти

KIMYO FANLARI/ CHEMICAL SCIENCES/ ХИМИЧЕСКИЕ НАУКИ

Axmedov Voxid Nizomovich
Buxoro Davlat Texnika Universiteti, professor
Umarov Baqo Bafoyevich
Buxoro Davlat Universiteti, professor
Kodirov Abduaxad Abduraximovich
Qarshi davlat Universiteti, professor
Qurbonov Mingniqul Jumagulovich
Qarshi davlat Universiteti, dotsent
Turemuratov Sharibay Naurizbaevich
O‘zbekiston Respublikasi Fanlar akademiyasi Qoraqalpog‘iston bo‘limi Bosh ilmiy kotibi, professor
Tursunov Kaxor Shonazarovich
Qarshi davlat Universiteti, professor
Do‘stov Hamro
Buxoro Davlat Texnika Universiteti, professor
Panoyev Nodir Shavkatovich
Buxoro Davlat Texnika Universiteti, dotsent.
Olimov Bobur Bahodir o‘g‘li
Buxoro Davlat Texnika Universiteti, dotsent.
Kasimov Sherzod Abduazimovich
Termiz Davlat Universiteti, professor.
Aliqulov Rustam Valiyevich
Termiz Davlat Universiteti, professor
Tropov Xamza Tursunovich
Samarqand Davlat Universiteti, professor

IQTISODIYOT FANLARI/ ECONOMICS/ ЭКОНОМИЧЕСКИЕ НАУКИ

Nizamov Aslitdin Badreddinovich
Buxoro Davlat Texnika Universiteti, professor.
Boboyev Akmal Choriyevich
Buxoro Davlat Texnika Universiteti, professor.
Ergashxojayeva Shaxnoza Djasurovna
Toshkent davlat iqtisodiyot universiteti, professor
Saidaxmedova Nodira Ilxomovna
Toshkent davlat iqtisodiyot universiteti, professor
Muxsinov Bekzod Toxirovich
Buxoro Davlat Texnika Universiteti, dotsent.
Shakarov Qulmat Ashirovich
Renessans ta'lim universiteti,dotsent.
Jumaeva Zulfiya Qayumovna
Buxoro Davlat Texnika Universiteti, dotsent.
Norova Salomat Yusupovna
Buxoro Davlat Texnika Universiteti, dotsent.
Avezova Shaxnoza Maxmudjonovna
Buxoro Davlat Texnika Universiteti, dotsent.
Jalilov Jamshid G‘anijonovich
Toshkent davlat iqtisodiyot universiteti, dotsent
Xasanova Gulruh Djumanazarovna
Buxoro Davlat Texnika Universiteti, dotsent.

PEDAGOGIKA FANLARI/ PEDAGOGICAL SCIENCES / ПЕДАГОГИЧЕСКИЕ НАУКИ

Siddiqova Sadoqat G‘afforovna
Buxoro Davlat Texnika Universiteti, dotsent.
Zoyirov Erkin Xalilovich
Buxoro Davlat Texnika Universiteti, DSc, professor
Gaffarov Feruz Hasanovich
Puchon Universiteti, *professor*
Ruziyev Davron Yuldashevich
Buxoro Davlat Universiteti, professor
Muradova Firuza Rashidovna
Buxoro Davlat Texnika Universiteti, professor.
Yunusova Gulandon Samiyevna
Buxoro Davlat Texnika Universiteti, professor.
To‘xtayeva Zebo Sharifovna
Buxoro Davlat Texnika Universiteti, professor.
Navro‘zova Gulchixra Nigmatovna
Buxoro Davlat Texnika Universiteti, dotsent.
Uzoqova Laylo Polvonovna
Buxoro muhandislik – texnologiya instituti, dotsent.
To‘raqulova Marjona Qiyom qizi
Buxoro Davlat Texnika Universiteti, dotsent.
Muradova Zarina Rashidovna
Buxoro Davlat Texnika Universiteti, dotsent.
Murodov Sanjar Aslonovich
Buxoro Davlat Texnika Universiteti, dotsent.
Shodiyev Jahongir Jo‘raqulovich
Buxoro Davlat Texnika Universiteti, dotsent
Gaffarov Laziz Hasanovich
Buxoro muhandislik – texnologiya instituti, dotsent.

Tahririyat a’zolarning to‘liq ro‘yxati jurnalning www.devos.uz saytida keltirilgan.

Tahririyat hay’ati a’zolarining tarkibini o‘zgartirish hamda qo‘shimcha kiritish yoki tahririyat a’zoligidan chiqarish bosh muharrir tomonidan amalga oshiriladi.

ISSN 3030-3907

Development of science ilmiy jurnali 2024 yil 15 aprel kuni №256445 sonli guvohnoma bilan Davlat ro‘yxatidan o‘tgan.

To‘plam elektron ko‘rinishda (pdf formatida) mualliflarga taqdim etiladi. To‘plamga kiritiladigan maqolalarning mazmuni, ulardagi statistik ma’lumotlar va me’yoriy hujjatlarning to‘g‘riligiga hamda har qanday fikr - mulohazalarga mualliflarning o‘zlari mas’ul hisoblanadilar. Talablarga javob bermaydigan maqolalar to‘plamga kiritilmaydi. Tashkiliy qo‘mita maqola matnini qisqartirish, qisman tuzatish kiritish, seksiyalarga joylashtirish huquqiga ega.

ISSN 3030-3907

UDK: 004.8, 801.8, 81, 81.32

LEKSIK SINONIMLAR OMBORINI WORD2VEC VA FASTTEXT MODELLARI YORDAMIDA HOSIL QILISH

Khajibaeva Surayyo Makhmudjonovna
Abu Rayhon Beruniy nomidagi Urganch davlat universiteti, o‘qituvchi
E-mail: surayyo.khajiboyeva@gmail.com

Anotatsiya. Ushbu maqolada mashinali o‘qitish algoritmlari yordamida o‘zbek tilining leksik sinonimlarni avtomatik ravishda ajratib olish muammosi keltirilgan. Faqat lug‘atga asoslangan leksik sinonimlarni avtomatik ravishda ajratib olish har doim ham to‘liq samarali emas, chunki lug‘atda "sinonimlar" sifatida keltirilgan so‘zlar haqiqiy matnda uchraydigan aniq kontekstga mos kelmasligi mumkin. Agglyutinativ tillar guruhiga kiruvchi o‘zbek tili uchun Word2Vec va FastText usullari asosida olingan natijalar UzWordNet sinsetlari asosida ajratib olindi va taqqoslandi. Tahlil natijalari shuni ko‘rsatdiki, FastText modeli natijalari Word2Vec modeli natijalariga nisbatan sezilarli darajada oshdi. Olingan natijalar shuni ko‘rsatadiki, agglyutinativ tuzilishga ega o‘zbek tili uchun FastText modeli asosida hosil bo‘lgan vektor maydoni sinonim qatorga kiritilgan ko‘plab so‘zlarning semantik yaqinligini muvaffaqiyatli aks ettiradi.

Kalit so‘zlar: so‘zlarni joylashtirish, Word2Vec, FastText, vektor o‘lchami, WordNet, sinset, maqsad so‘z

Annotation. This paper addresses the issue of automatically extracting lexical synonyms in the Uzbek language using machine learning algorithms. Extracting lexical synonyms automatically based solely on a dictionary is not always fully effective, since words listed as 'synonyms' in a dictionary may not correspond to their exact usage in real text. Results for the Uzbek language (which belongs to the agglutinative language group) were obtained using Word2Vec and FastText and extracted from the UzWordNet synsets, which were then compared. Analysis showed that the FastText model achieved significantly higher performance than the Word2Vec model. These results demonstrate that the vector space constructed using the FastText model for Uzbek accurately reflects the semantic proximity of many words in the synonymous series, despite the language's agglutinative structure.

Keywords: word embedding, Wor2Vec, FastText, vector size, WordNet, synset, target words.

Аннотация. В данной работе рассматривается проблема автоматического извлечения лексических синонимов в узбекском языке с использованием алгоритмов машинного обучения. Автоматическое извлечение лексических синонимов исключительно на основе словаря не всегда полностью

ISSN 3030-3907

эффективно, поскольку слова, перечисленные в словаре как «синонимы», могут не соответствовать их точному употреблению в реальном тексте. Результаты для узбекского языка (который относится к агглютинативной языковой группе) были получены с использованием Word2Vec и FastText и извлечены из наборов синонимов UzWordNet, которые затем были сравнены. Анализ показал, что модель FastText продемонстрировала значительно более высокую производительность, чем модель Word2Vec. Эти результаты показывают, что векторное пространство, построенное с использованием модели FastText для узбекского языка, точно отражает семантическую близость многих слов в синонимичных рядах, несмотря на агглютинативную структуру языка.

Ключевые слова: векторное представление слов, Wor2Vec, FastText, размер вектора, WordNet, синсет, целевые слова.

KIRISH.

Sinonimlar - bu ma'lum bir kontekstda bir xil yoki deyarli bir xil ma'noga ega bo‘lgan so‘zlar yoki iboralar. Ular har doim ham har bir vaziyatda bir-biriga mos kelavermaydi, chunki uslub, hissiy rang, qo‘llanilish yoki ma'no darajasida farqlar mavjud bo‘lishi mumkin. Masalan, ikkita so‘z o‘xshash tushunchani anglatishi mumkin, ammo biri rasmiy, ikkinchisi esa norasmiy bo‘lishi mumkin.

Tilshunoslikda sinonimlarning bir nechta turlari mavjud. Mutlaq sinonimlar - bu barcha kontekstlarda ma'no o‘zgarmasdan bir-birini almashtira oladigan so‘zlar, garchi bunday holatlar tabiiy tillarda juda kam uchraydi. Qisman yoki yaqin sinonimlar ko‘proq uchraydi; ular markaziy ma'noga ega, ammo nuans yoki qo‘llanilish jihatidan biroz farq qiladi. Stilistik sinonimlar bir xil asosiy g‘oyani ifodalaydi, lekin rasmiy, neytral yoki so‘zlashuv tili kabi turli xil nutq uslublariga tegishli. Semantik yoki ideografik sinonimlar ma'no tuslari, intensivligi yoki kontseptual urg‘usi bilan farq qiladi.

Leksik sinonimlar - bu nutqning bir qismiga tegishli bo‘lgan va lug‘at darajasida ma'no jihatidan yaqin yoki bir xil bo‘lgan so‘zlar. Boshqacha qilib aytganda, ular bir xil obyekt, harakat, sifat yoki holatni ma'lum darajada semantik o‘xshashlik bilan ifodalashi mumkin bo‘lgan alohida leksik birliklardir. Masalan, ikkita ot, ikkita fe'l yoki ikkita sifat, agar ular bir-biriga chambarchas bog‘liq ma'nolarni ifodalasa, leksik sinonimlar deb hisoblanishi mumkin. Biroq, leksik sinonimlar odatda ma'no, chastota, stilistik qiymat yoki foydalanish doirasi jihatidan farq qiladi.

Leksik sinonimlarni o‘rganish tilshunoslikda muhim ahamiyatga ega, chunki u tilning boyligi va moslashuvchanligini ochib berishga yordam beradi. Shuningdek, u leksikografiya, til o‘qitish, matnni soddalashtirish, parafrazalash,

ISSN 3030-3907

mashina tarjimasi, semantik qidiruv va boshqa tabiiy tilni qayta ishlash vazifalarida juda foydali. Shu ma'noda, leksik sinonimiya nafaqat nazariy tushuncha, balki amaliy lingvistik resurs hamdir.

An'anaviy tilshunoslikda sinonimlar asosan lug‘atlar asosida tavsiflanadi. Turli sinonim lug‘atlarida so‘zga ma'no jihatidan yaqin deb hisoblangan boshqa birliklar beriladi va ular til tizimida sinonim qatorlarni hosil qiladi. Biroq, amaliy tajriba shuni ko‘rsatadiki, faqat lug‘at ma'lumotlariga asoslangan sinonimlarni avtomatik ravishda aniqlash har doim ham samarali natijalar bermaydi. Buning sababi, lug‘atda sinonim sifatida berilgan birliklar har doim ham haqiqiy matn kontekstida bir-birini to‘liq almashtira olmaydi. Ba'zi hollarda, ular ma'no nozikligi, stilistik farqlar yoki semantik ma'nolar tufayli turli vazifalarda qo‘llaniladi. Masalan, "yuz" va "bashara" lug‘atda bir xil sinonimlar guruhiga kiritilgan. Biroq, "yuz" ijobiy ma'nolarda, "bashara" esa salbiy ma'nolarda ishlatiladi. Natijada, lug‘atga asoslangan mexanik almashtirish matn mazmuniga yoki uslubiga zarar yetkazishi mumkin.

So‘nggi yillarda tabiiy tilni qayta ishlashda mashina o‘rganish usullarining keng qo‘llanilishi o‘zbek so‘zlarining leksik semantikasini o‘rganish uchun yangi imkoniyatlar yaratdi. Xususan, so‘zlarni joylashtirish deb nomlanuvchi usullar matematik makonda so‘zlarning semantik o‘xshashligini modellashtirish imkonini beradi. Bunday modellarda har bir so‘z ko‘p o‘lchovli vektor sifatida ifodalanadi va o‘xshash kontekstlarda ishlatiladigan so‘zlar vektor makonida bir-biriga yaqin joylashgan. Natijada, katta matnli korpuslar asosida so‘zlar orasidagi yashirin semantik munosabatlarni aniqlash mumkin.

O‘zbek tili mashina o‘rganish modellari yordamida katta korpuslarda chuqur o‘rganilmagan. Bunday modellarning eng mashhur vakillaridan biri Word2Vec [1]. Ushbu model kontekst asosida so‘zlarning vektor tasvirlarini o‘rganadi va semantik jihatdan yaqin birliklarni aniqlash imkonini beradi. Keyinchalik bu yondashuv FastText modeli bilan yanada rivojlantirildi [2]. FastText modeli so‘zlarni nafaqat butun birliklar, balki pastki so‘z segmentlari sifatida ham tahlil qiladi. Bu xususiyat, ayniqsa, agglyutinativ tillar uchun muhimdir, chunki bunday tillarda so‘z shakllari ko‘plab morfologik qo‘shimchalar orqali hosil bo‘ladi. O‘zbek tili ham agglyutinativ til bo‘lgani uchun uning morfologik boyligi so‘z vektorlarini o‘rganishga ixtisoslashgan yondashuvlarni taqozo etadi.

ASOSIY QISM. So‘zlarni joylashtirish(word embadding) – matndagi so‘zlarning ma’nosini saqlashga harakat qiladigan n o‘lchovli taqsimlangan tasvir [3]. Ushbu usullar yordamida so‘zlar k-o‘lchovli vektor ko‘rinishda tasvirlanadi. Ma’lumki, so‘zlar joylashtirish usullari kattaroq korpuslar ustida o‘qitilganda yaxshi natija beradi. Masalan, Google News ma’lumotlarining bir qismi asosida

ISSN 3030-3907

o‘qitilgan Word2Vec modeli taxminan 100 milliard so‘zni o‘z ichiga qamrab olgan. Mazkur tadqiqot o‘zbek tili uchun Wikipedia o‘zbek matnlari, daryo.uz va shu kabi bir nechta elektron manbalar hamda maktab korpusi [4] asosida qilindi.Tadqiqot doirasida yuqorida keltirilgan korpusdan foydalanib, O‘zbek tilining lotin alifbosidagi matnlari asosida Word2Vec va FastText so‘zlarni joylashtirish usullari o‘qitildi. Keyinchalik ushbu modellar natijalari o‘zaro solishtirilib, ularning leksik sinonimlarni aniqlashdagi imkoniyatlari qiyosiy baholandi. WordNet[5] asosida boshlang‘ich tadqiqot uchun 1549 ta maqsad so‘zga tegishli synsetlar ajratib olindi va synset tarkibidagi leksik birliklar sinonimik qator sifatida ko‘rib chiqildi. Shundan so‘ng, barcha ma’lumotlar “so‘z — synset — sinonimlar” ko‘rinishida yagona faylga jamlandi.

Word2Vec modeli matndagi har bir unikal token uchun zich vektorli ifoda hosil qiluvchi neyron tarmoqqa asoslangan yondashuv hisoblanadi. Ushbu model so‘zlar orasidagi semantik munosabatlarni ularning matndagi birgalikda uchrashish xususiyatlari orqali o‘rganadi. Vektorlar fazosida bir-biriga yaqin joylashgan so‘zlar, xususan kosinus o‘xshashligi qiymati 1 ga yaqin bo‘lgan birliklar, ma’no jihatidan o‘xshash deb talqin etiladi. Shuning uchun Word2Vec sinonimlarni aniqlash, mashina tarjimasi, avtomatik teglash va boshqa NLP vazifalarida samarali vosita sifatida qo‘llanadi.

Word2Vec modeli tajriba jarayonida vector_size=200, window=3, min_count=3, sg=1, negative=15, sample=1e-4, epochs=10, workers=8 va seed=42 parametrlaridan foydalangan holda o‘qitildi. Bunda sg=1 qiymati modelning Skip-gram usulida ishlashini bildiradi. Ushbu parametrlar to‘plami o‘zbek tilidagi leksik birliklarning kontekstual-semantik xususiyatlarini vektor fazosida ifodalash hamda sinonimik yaqinliklarni aniqlash uchun tanlandi.

FastText modeli matndagi leksik birliklarning vektorli ifodalarini hosil qiluvchi va ularning semantik munosabatlarini o‘rganuvchi taqsimlangan tasvirlash usullaridan biridir. Modelning asosiy xususiyati shundaki, u har bir so‘zni belgilar darajasidagi n-grammalar yig‘indisi orqali ifodalaydi. Bu esa kam uchraydigan yoki turli morfologik shakllarda keladigan so‘zlar uchun ham barqarorroq vektorlar hosil qilish imkonini beradi. Shu bois FastText o‘zbek tilidagi sinonimik munosabatlarni aniqlashda samarali model sifatida qaraladi.

FastText modeli tajriba jarayonida vector_size=300, window=7, min_count=3,sg=1, negative=15, sample=1e-4, min_n=3, max_n=6 va workers=8 parametrlaridan foydalangan holda train qilindi. Ushbu sozlamalarda model Skip-gram mexanizmi asosida ishlaydi hamda har bir so‘zni belgilar darajasidagi n-grammalar orqali ham ifodalaydi. Bu tadqiqotda FastText modelida min_n=3 va max_n=6 parametrlaridan foydalanildi, ya’ni so‘zlarning 3 belgili, 4 belgili, 5

ISSN 3030-3907

belgili va 6 belgili belgilar ketma-ketliklari (character n-grams) hisobga olindi. Ushbu yondashuv o‘zbek tilidagi morfologik o‘zgarishlarni, ayniqsa qo‘shimchalanish hodisalarini, samaraliroq aks ettirish imkonini berdi. Bu yondashuv, ayniqsa, o‘zbek tilidagi qo‘shimchalanish hodisalari kuchli bo‘lgan morfologik shakllarni yaxshiroq qamrab olish imkonini beradi. Shu sababli FastText modeli leksik-semantik yaqinliklarni, jumladan sinonimik munosabatlarni aniqlashda muhim vosita sifatida tanlandi.

Sinonim kandidatlarni ajratib olish jarayonida dastlab har bir target so‘z uchun embedding fazodan eng yaqin qo‘shni birliklar tanlab olindi. Yaqinlik darajasi so‘z vektorlari orasidagi kosinus o‘xshashligi asosida baholandi va tajribalarda top-k qo‘shnilar soni mos ravishda 10 va 50 qiymatlarida sinab ko‘rildi. Turli k qiymatlarining qo‘llanishi o‘zbek tilining agglyutinativ xususiyati bilan izohlanadi, chunki bunday tillarda bitta leksik birlik ko‘plab yuzaki token shakllarida namoyon bo‘lishi mumkin.

Tajriba davomida kichik top-k qiymatlari qo‘llanganda yaqin qo‘shnilar ro‘yxatida ko‘pincha maqsad so‘zning morfologik variantlari ustunlik qilishi kuzatildi. Masalan, abadiy so‘zi uchun dastlabki eng yaqin qo‘shnilar abadiylik, abadiydir, abadiyati, abadiyat, abadiylikni, abadiyatga, abadiyligi, abadiyatning, abadiylikka, abadiyligini, abadiylar, abadiyligiga, abadiylashtiradi, abadiylashtirilgan kabi ko‘plab so‘zlardan tashkil topgan. Bunday holat semantik jihatdan yaqin, ammo shaklan boshqa sinonim birliklarning qo‘shnilar ro‘yxatida pastroq o‘rin egallashiga sabab bo‘ladi. Natijada tor qamrovli top-k tanlovlari sinonimik qatorning to‘liqroq tiklanishiga to‘sqinlik qiladi. Shuning uchun, FastText modelidan foydalanganda agglyutinativ tillar uchun sinonim qatorini to‘liq tiklash uchun 10 dan yuqori k qiymatini tanlash maqsadga muvofiqligi aniqlandi. Umuman olganda, natijalar shuni ko‘rsatadiki, FastText va WordNet integratsiyasiga asoslangan yondashuv o‘zbek tilida sinonim resurslarni yaratishda samarali hisoblanadi.

Ushbu tadqiqotning cheklovlaridan biri shundaki, tajriba nisbatan cheklangan maqsadli so‘zlar to‘plamida o‘tkazildi. Kelajakda ushbu metodologiyani kengroq elektron lug‘at va kengroq leksik korpusga qo‘llash orqali sinonim resursini yanada boyitish bo‘yicha tadqiqotlar rejalashtirilgan.

Ushbu tadqiqotda yaqin qo‘shnilar sonini oshirish strategiyasi qo‘llanildi. Kattaroq top-k qiymatlaridan foydalanish morfologik variantlar bilan cheklanib qolmasdan, semantik jihatdan relevant bo‘lgan boshqa leksik birliklarni ham qamrab olish imkonini berdi. Keyingi bosqichda so‘zlarni joylashtirish fazodan olingan yaqin qo‘shnilar WordNet asosida oldindan shakllantirilgan sinonim

ISSN 3030-3907

to‘plamlar bilan taqqoslandi. Qo‘shnilar ro‘yxati bilan WordNet sinonimlari orasidagi kesishma qismi sinonim sifatida qabul qilindi.

1-Jadval.

Word2Vec va FastText modellari yordamida 10, 50 ta eng yaqin qo‘shnilar olinganda aniqlik darajasi va lug‘atdan tashqari so‘zlar(OOV) miqdori(%).

Model	Aniqlik@10	Aniqlik@ 50	OOV (%)
Word2Vec	0.04532	0.014564	34%
FastText	0.027179	0.01472	16.5%

1-jadvaldan ko‘rish mumkinki, Word2Vec modeli 10 ta yaqin qo‘shnilarga nisbatan yaxshiroq aniqlik darajasini ko‘rsatdi. Bundan ko‘rinadiki, kamroq eng yaqin qo‘shnilar olinganda Word2Vec aniqlik ko‘rsatkichi yuqori bo‘lsa ham ko‘proq yaqin qo‘shnilar olinganda FastText ustunlik qiladi. Bundan tashqari, FastTextning OOV lar miqdori ham kamroq ekanligini ko‘rishimiz mumkin.

Olingan sinonimlar to‘plamlari o‘zbek tabiiy tilini qayta ishlashning bir qancha muhim sohalarida qo‘llanilishi mumkin. Birinchidan, ular sinonim lug‘atlari va o‘zbek tili uchun semantik ma'lumotlar bazalari kabi leksik resurslarni kengaytirish uchun ishlatilishi mumkin. Ikkinchidan, bu natijalar axborotni qidirish va semantik qidiruv tizimlari uchun qimmatlidir, bunda sinonimlarni kengaytirish foydalanuvchi so‘rovlarini turli leksik shakllar orqali ifodalangan tegishli hujjatlar bilan moslashtirishga yordam beradi. Uchinchidan, ular murakkab yoki unikal so‘zlarni semantik jihatdan ekvivalent va tanishroq alternativalar bilan almashtirish orqali matnni soddalashtirish va ta'lim dasturlarini qo‘llab-quvvatlashi mumkin. To‘rtinchidan, sinonimlar resurslari mashina tarjimasi va parafraza yaratishda foydalidir, bu yerda leksik o‘zgarishlar asosiy ma'noni o‘zgartirmasdan ishlov berilishi kerak. Bundan tashqari, ular leksik siyraklikni kamaytirish va semantik jihatdan yaqin birliklarni guruhlash orqali matn tasnifini, klasterlashni va mavzu modellashtirishni yaxshilashi mumkin. Bunday sinonimlar to‘plamlari savol-javob tizimlari, chatbotlar va dialog tizimlari uchun ham muhimdir, chunki ular foydalanuvchi kiritishi va saqlangan bilimlar o‘rtasidagi semantik moslikni kuchaytiradi. Shuning uchun olingan natijalar nafaqat lingvistik jihatdan mazmunli, balki mustahkam va semantik jihatdan xabardor o‘zbek tili texnologiyalarini yaratish uchun ham amaliy ahamiyatga ega.

XULOSA.

Ushbu tadqiqotda Word2Vec va FastText modellarining o‘zbek tilida leksik sinonimlar resursini qurish vazifasidagi samaradorligi taqqoslandi. Natijalar FastText modelining ustunligini ko‘rsatdi: u sinonimik qatordagi ko‘proq birliklarni qamrab oldi va Word2Vecga nisbatan kamroq OOV holatlarini yuzaga keltirdi. Bu holat FastTextning character n-grams asosidagi subword modeling

ISSN 3030-3907

yondashuvi bilan izohlanadi, chunki bunday mexanizm o‘zbek tilining agglyutinativ va morfologik jihatdan boy xususiyatlarini yaxshiroq aks ettiradi.

Shu bilan birga, embedding fazodan olingan eng yaqin qo‘shnilar ro‘yxatidagi barcha birliklar sinonim bo‘lavermaydi. Ular orasida gipernim, antonim, omonim, ham mavjud bo‘lishi mumkin. Shu sababli WordNet ushbu kandidatlarni leksik-semantik nuqtai nazardan tekshiruvchi qatlam sifatida xizmat qildi. So‘zning vector koordinatalari yordamida topilgan yaqin qo‘shnilar WordNet asosida tayyorlangan sinonimik to‘plamlar bilan solishtirildi va kesishgan qism sinonim sifatida qabul qilindi.

Ayrim maqsad so‘zlarning so‘zlarni joylashtirish lug‘atida topilmagani esa WordNetdagi birliklarning lemma shaklida, vektor fazodagi birliklarning esa asosan token shaklida ifodalanganligi bilan izohlanadi. Ushbu lemma-token nomuvofiqligi OOV holatlar sonining oshishiga olib keldi. Umuman olganda, natijalar FastText va WordNet integratsiyasiga asoslangan yondashuv o‘zbek tilida sinonimik resurslarni shakllantirishda samarali ekanini ko‘rsatadi.

FOYDALANILGAN ADABIYOTLAR RO‘YXATI

1. T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient Estimation of Word Representations in Vector Space,” in Proceedings of the 1st International Conference on Learning Representations (ICLR Workshop), Scottsdale, AZ, USA, 2013.

2. P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching Word Vectors with Subword Information,” in Proc. 5th International Conference on Learning Representations (ICLR), Toulon, France, 2017.

3. Elov, B. (2024). Matnli ma’lumotlarni Word2Vec, GloVe va FastText chuqur o‘rganish usullari vositasida qayta ishlash. Raqamli Transformatsiya va Sun’iy Intellekt ilmiy jurnali, 2(5), 225.

4. K. Madatov and S. Khajibaeva, “Extraction and Data Analysis Basis Words: Case Study on School Corpus,” in 2025 IEEE XVII International Scientific and Technical Conference on Actual Problems of Electronic Instrument Engineering (APEIE), 2025, pp. 1–5.

5. N. Abdurahmonova, “UzWordNet: A Lexical-Semantic Database for the Uzbek Language,” in Proceedings of the 10th Global WordNet Conference (GWC), Wroclaw, Poland, 2019, pp. 8–15.

ISSN 3030-3907