

ISSN 2181-922X

LANGUAGE & CULTURE

UZBEKISTON O'ZBEKISTON

TIL VA MADANIYAT

KOMPYUTER
LINGVISTIKASI

2024 Vol. 3 (6)

www.compling.tsuull.uz

ISSN 2181-922X

O‘ZBEKISTON

TIL VA MADANIYAT

KOMPYUTER LINGVISTIKASI

2024 Vol. 3 (6)

compling.tsuull.uz

Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti

Bosh muharrir:

Botir Elov

Bosh muharrir o'rinbosari:

Shahlo Hamroyeva

Mas'ul kotib:

Oqila Abdullayeva

Tahrir kengashi

Shuhrat Sirojiddinov (O'zbekiston), Eshref Adali (Turkiya), [Viktor Zaxarov] (Rossiya), Vladimir Benko (Slovakiya), Ayrat Gatiatullin (Tataristan), Rinat Gilmullin (Tataristan), Murat O'rxun (Turkiya), Suyun Karimov (O'zbekiston), Abduvali Qarshiyev (O'zbekiston), Muxammadjon Musayev (O'zbekiston), Kamoliddin Shukurov (O'zbekiston), O'tkir Hamdamov (O'zbekiston), Tal'at Zuparov (O'zbekiston), Bahodir Mo'minov (O'zbekiston), Faxriddin Nurullayev (O'zbekiston), Zulxumor Xolmanova (O'zbekiston), Muqaddas Abdurahmonova (O'zbekiston), Elova Dilrabo (O'zbekiston), Ruhillo Alayev (O'zbekiston), Rasuljon Atamuratov (O'zbekiston), Malika Abdullayeva (O'zbekiston), Mannon Ochilov (O'zbekiston), Xolisa Axmedova (O'zbekiston), Zilola Xusainova (O'zbekiston), Uldona Abdurahmonova (O'zbekiston).

Jurnal haqida ma'lumot

“O'zbekiston: til va madaniyat. Kompyuter lingvistikasi” seriyasi – Oliy attestatsiya komissiyasi ilmiy nashrlar ro'yxatidagi “O'zbekiston: til va madaniyat” akademik jurnalining ilovasi hisoblanib, unda professor-o'qituvchilar, doktorantlar, stajor-tadqiqotchilar, mustaqil izlanuvchilar, magistrantlarning kompyuter lingvistikasi, jumladan, tabiiy tilga ishlov berish (NLP), o'zbek tilining formal grammatikasi, korpus lingvistikasi, mashina tarjimasi, nutqni qayta ishlash tizimlari, intellektual tizimlar, kompyuter leksikografiyasi hamda lingvistik ontologiyalar kabi sohalarga oid tadqiqotlari nashr qilinadi.

Jurnal ilovasi bir yilda to'rt marta chop etiladi.

O'zbek, turk, rus va ingliz tillarida yozilgan maqolalar qabul qilinadi.

Jurnalda kitoblarga yozilgan taqrizlar, adabiyotlar sharhi, konferensiyalar hisobotlari va tadqiqot loyihalari natijalari ham e'lon qilinadi.

Mualliflar fikri tahririyat nuqtayi nazaridan farq qilishi mumkin.

“O'zbekiston: til va madaniyat. Kompyuter lingvistikasi” seriyasi 2023-yildan chiqa boshlagan.

Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti. O'zbekiston, Toshkent, Yakkasaroy tumani, Yusuf Xos Hojib ko'chasi, 103-uy.

E-mail: kompling@navoiy-uni.uz

Website: kompling.tsuull.uz

Alisher Navo'i Tashkent State University of the Uzbek Language and Literature

Chief editor:	Botir Elov
Deputy editor-in-chief:	Shahlo Hamroyeva
Responsible secretary:	Oqila Abdullayeva

Editorial board

Shukhrat Sirojiddinov (Uzbekiston), Eshref Adali (Turkiye), [Viktor Zakharov] (Russia), Vladimir Benko (Slovakia), Ayrat Gatiatullin (Tataristan), Rinat Gilmullin (Tataristan), Murat Orhun (Turkey), Suyun Karimov (Uzbekistan), Abduvali Karshiyev (Uzbekistan), Mukhammadjon Musayev (Uzbekistan), Kamoliddin Shukurov (Uzbekistan), O'tkir Hamdamov (Uzbekistan), Tal'at Zuparov (Uzbekistan), Bahadir Mo'minov (Uzbekistan), Fakhridin Nurullayev (Uzbekistan), Zulkhumor Kholmanova (Uzbekistan), Muqaddas Abdurakhmonova (Uzbekistan), Elova Dilrabo (Uzbekistan), Ruhillo Alayev (Uzbekistan), Rasuljon Atamuratov (Uzbekistan), Malika Abdullayeva (Uzbekistan), Mannon Ochilov (Uzbekistan), Kholisa Akhmedova (Uzbekistan), Zilola Khusainova (Uzbekistan), Uldona Abdurakhmonova (Uzbekistan).

Information about the magazine

"Uzbekistan: language and culture. "Computer Linguistics" series is an appendix of the academic journal "Uzbekistan: Language and Culture" in the list of scientific publications of the Higher Attestation Commission, in which computer linguistics, including natural language processing (NLP) of professors-teachers, doctoral students, intern-researchers, independent researchers, master's students, researches related to formal grammar of the Uzbek language, corpus linguistics, machine translation, speech processing systems, intelligent systems, computer lexicography and linguistic ontologies are published.

The magazine supplement is published four times a year.

Articles written in Uzbek, Turkish, Russian and English languages are accepted.

The journal also publishes book reviews, literature reviews, conference reports, and research project results.

The opinion of the authors may differ from the editorial point of view.

"Uzbekistan: language and culture. "Computer Linguistics" series has been published since 2023.

Tashkent State University of Uzbek Language and Literature named after Alisher Navoi. Yusuf Khos Hajib street, 103, Yakkasaray district, Tashkent, Uzbekistan.

E-mail: kompling@navoiy-uni.uz

Website: kompling.tsuull.uz

MUNDARIJA

Talat Zuparov

Word2vec metodi orqali matnlarni raqamlashtirish va mashinali o'qitish usullari orqali qayta ishlash.....6

Umidjon Yodgorov

O'zbek tili frazemalarining morfologik shakli va variantlari tadqiqiga an'anaviy va korpus tahlili yondashuvi.....29

Go'zal Erkinjonova

Fe'l so'z turkumiga oid birliklarning leksik-grammatik xususiyati.....49

Botir Elov, Maftuna Baratova

Pos (part of speech) teglash usullari.....62

Zebo Qodirova

Tibbiy atamalarni tur va sinflarga ajratishda ontologik tamoyillar.....78

Gulira'no Nuriddinova

Tabiiy tilni qayta ishlashda eganing modellari.....90

Botir Elov, Maftuna Ahmedova

N-gramlar asosida imloni tuzatish tizimini ishlab chiqish.....101

Mavluda Urazaliyeva

Audiomatnlarni korpusga kiritish muammolari tahlili.....115

Gulruksor Xoliyorova

Leksikografik manbalarda tezauruslar tadqiqi.....126

CONTENT

Talat Zuparov

Digitization of texts using the word2vec method and processing through machine learning techniques.....27

Umidjon Yodgorov

The traditional and corpus-based approach to the study of morphological forms and variations of uzbek phraseology.....46

Go'zal Erkinjonova

Lexical and grammatical characteristics of verbal word class units.....60

Botir Elov, Maftuna Baratova

Pos (part of speech) tagging methods.....76

Zebo Kodirova

Ontological principles in the division of medical terms into types and classes.....88

Gulira'no Nuriddinova

Models of subject in natural language processing.....100

Botir Elov, Maftuna Ahmedova

Development of a spell correction system based on n-grams.....113

Mavluda Urazaliyeva

Analysis of problems in incorporating audio texts into a corpus.....124

Gulruxsor Xoliyorova

Research on thesauri in lexicographic sources.....134

LEKSIKOGRAFIK MANBALARDA TEZAURUSLAR TADQIQI

Gulruxsor Xoliyorova¹

Annotatsiya. Mazkur maqolada tezauruslar, ularning lug'atlar yaratishdagi ahamiyati hamda tasnif qilish tamoyillari haqida gap boradi.

Tezaurus lug'atlarining afzalliklarini misollar yordamida izohlanadi. Tezaurus lug'atlari, so'zlar va ularning o'zaro aloqalarini tahlil qilish uchun yaratilgan ma'lumotlar bazasi hisoblanadi. Ular bir-biriga yaqinligi bo'lgan so'zlar to'plamidir. Bu lug'atlarning asosiy maqsadi, bir so'zning ma'nosi bo'yicha boshqa so'zlarni topish va ularga o'xshash so'zlar topishdir. Bu matnlar yoki turli xil ma'lumotlar bazalarida ishlatilishi mumkin. Tezaurus lug'atlari, matnlarni tahlil qilishda va tarjimada yordam berishda keng ishlatiladi. Ular matnda o'xshash yoki o'zaro bog'liq so'zlar ro'yxatini ko'rsatadi. Bu matnda qidirayotgan so'zni topishga yordam beradi. Shuningdek, tezaurus lug'atlari, matnlarni taqsonlarga bo'lishda va taqsonlarni shakllantirishda ishlatiladi. Tezaurus lug'atlari, boshqa lug'atlarga nisbatan bir necha foydali imkoniyatlar taqdim etadi. Ular so'zlarning ma'nolarini tahlil qilish, ularning o'zaro aloqalarini aniqlash, so'zlar va ularning aloqadorligini tushuntirish va matnlarni taqsonlarga bo'lishda yordam beradi. Bu, matnlar va turli ma'lumotlar bazalaridagi ishlarimizda juda foydali bo'ladi. Tezaurus lug'atlari, dastlabki qarashda, lug'atlar bilan bir xil deb o'ylansa ham, o'zaro farqlanuvchi xususiyatlarga ega.

Kalit so'zlar: *tezaurus, lug'at, semantik bog'lanish, WordNet.*

Kirish

Leksikografik tadqiqotlar sohasida tezauruslar va WordNet tizimlari yaratish bo'yicha keng qamrovli izlanishlar olib borilgan. Bu tizimlar so'zlarning ma'nolarini va ularning o'zaro munosabatlarini tartibga solishda muhim ahamiyat kasb etadi. Manbalarda qayd

¹Xoliyorova Gulruxsor G'ayratovna - O'zbekiston Milliy universiteti, Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasida o'qituvchisi.

E-pochta: guliholiyorova@gmail.com

etilishicha [Miller va boshq.], WordNet boshqa lug'atlardan farqli ravishda besh kategoriyaga ko'ra bo'linadi. Odatda an'anaiy lug'atlarda ot, fe'l, sifat, ravish va funksional so'zlar (yordamchi so'zlar) mavjud bo'lsa, WordNetda faqat shulardan to'rttasi, ya'ni ot, fe'l, sifat va ravish so'z turkumiga tegishli birliklar kiritiladi. WordNetga kiritiladigan so'zlar maxsus sinsetlarga biriktirilib, yordamchi so'z turkumiga tegishli birliklar tilning sintaktik komponenti sifatida qaralgan.

Asosiy qism

Wordnetda ot so'z turkumiga kiruvchi birliklar leksik xotirada mavzuiy iyerarxiya sifatida asos sanaladi. Fe'l so'z turkumidagi birliklar otga tobelanuvchi sifatida, sifat va ravish esa har ikki ot va fe'l uchun belgini ifodalovchi parametr sanaladi. Har bir leksik struktura turli tasnifiy hodisani ifodalashi, barcha sintaktik kategoriyalar uchun har bir alohida tashkil qilingan tamoyil leksik bilimlarni psixologik murakkablashtishda noto'g'ri ifodalanishi mumkinligi Miller va boshqalar tomonidan olib borilgan kuzatishlarda qayd etiladi. WordNetning o'ziga xos jihati shundaki, leksemalarning so'z shakliga ko'ra emas, balki ma'nosiga qarab leksik ma'lumotlar loyihalashtiriladi. Shu jihatdan WordNet an'anviy lug'atlardan ham ko'ra tezaurusga o'xshaydi.

Ta'kidlash o'rinliki, WordNetni yaratilishida uch turdagi lug'atlarning, ya'ni Laurence Urdang's revision of Rodale's The Synonym Finder (1978) and Robert L.Chapman's revision of Roget's International Thesaurus (1977) kabi manbalardan unumli foydalanilganligi qayd etiladi. Manbaga ko'ra, yuqorida nomlari qayd etilgan lug'atlar talabga to'la mos kelmagan. Muammo shundan iborat bo'lganki, tezaurusga kiritilgan birliklar alifbo tartibda bo'lgan: agar W_x bilan W_y so'z bir-biri bilan sinonim bo'lsa, ularning juftliklarini ikki marta alohida kiritish lozim bo'lgan, shunda W_x uchun alifbo tartib hamda W_y uchun boshqa alifbo tartibi tizimlashtirilgan. Mavzulashtirilgan tezaurusda ikki marta birliklarni qarab chiqish ehtiyoji bo'lgan. Dastlab alifbo tartibdagi so'zlar ro'yxati, keyin tezaurusdagi mos variant tanlangan. Bu kabi leksikografik yondashuv inson mehnatini ko'proq talab qiladi.

So'zlarning polisemiyasi ham giperonim va giponim darajada so'zning asosiga bog'liq bo'ladi, shu bois konseptlarga bog'lanish darajasi ham qaysidir ma'noda yaqin bo'ladi. Shuni ta'kidlash o'rinliki, tezaurusga berilgan barcha ta'riflarda uning tilning tizim ekanligiga urg'u beriladi. Buni shunday tushuntirish mumkin: har bir hodisa yoki tushuncha qaysidir ong va ongostida mavjud

bo'ladi. Kommunikatsiya jarayonida o'sha tizimlilik insonga matnni tez va to'g'ri tuzishga yordam beradi. Shu bois S.Osokina [Osokina, 2015] tezaurusning bilimlar tizimi ekanligi, tizimning umumiy mohiyatidan kelib chiqib, elementlar va ular o'zaro munosabatidan tashkil etishini asosli ta'kidlaydi. Elementlar semiotik ahamiyat kasb etib, material va semantik darajada ular o'rtasidagi munosabatlar mavjudligini asoslashga harakat qiladi. Elementlar o'rtasidagi bog'lanishni mavjud tizimlar strukturasi sifatida o'rganish asosli ekanligi e'tirof etiladi. Shu bilan bir qatorda ochiq tizim bo'lib, uning umumiy belgilarning hammasi ochiq tizim sanalishini, uning javob reaksiyasi xotira tizimida saqlanuvchi aniq strukturalar bilan bog'liq ravishda tashqi muhit bilan qanday munosabatga kirishuvi bilan bog'liq bo'lishini o'zining tezaurusga oid izlanishlarida dalillaydi.

Rus tili uchun RuTez lug'ati 1998-yili ijtimoiy-siyosiy terminlarining tezaurusi loyihasi asosida yaratila boshlangan [Aleksev va boshq., 2013]. Mualliflarning qayd etishicha, ushbu lug'at matnni avtomatik qayta ishlash hamda axborot qidiruv tizimida til va olam haqidagi bilimlarning manbasi sifatida foydalanishga mo'ljallangan. Unga ko'ra RuTez kompyuter resurslarni qayta ishlashda uch asosiy modellashtirish tamoyilidan foydalanilgan:

- an'anaviy axborot qidiruv tezaurusini qayta ishlash tamoyillari;

- WordNet tipidagi lingvistik resurslarni qayta ishlash an'analari;

- Formal ontologiyani yaratish tamoyillari

Olimning fikricha, matnni qayta ishlash orqali lingvistik resurslardan foydalanishda WordNet tipidagi lingvistik resurslarni qayta ishlash metodologiyasidan foydalanish zarur bo'ladi:

- RuTez tezaurusidagi birliklar tilda mavjud bo'lgan real ma'nolar asosida yaratilgan;

- ko'p bosqichli iyerarxik tuzilma tushunchalarning leksik-terminologik tizimi sanaladi;

- ko'p ma'noli so'z va ifodalarning ma'nosini izohlash tamoyillariga amal qilinadi.

Bizningcha ham, tezaurus nafaqat sinonimlardan iborat tuzilma, balki u olam haqidagi lisoniy ifodalarni konseptlar asosida umumlashtiradi. Bu borada yuqorida nomi qayd etilgan asarda quyidagi ilmiy mulohaza beriladi: "RuTez tezaurusining birliklari shu bilan birga nafaqat sinonimlar majmuasi, balki ontologik tadqiqotlarda iyerarxik tashkillashtirilgan tushunchalar tizimidagi unikal ism hamda unikal mohiyatga ega tushunchani ifodalaydi. Shu

bilan birga, tushunchalar o'rtasidagi munosabatlarni tavsiflashda uning formal mohiyati va askioma bilan munosabatlaridan foydalaniladi".

Tezaurus yondashuvi bo'yicha turlicha talqinlar ifodalangan. A.Juravlev bu borada quyidagi fikrlarni keltirib o'tadi [Juravlev, 2021]:

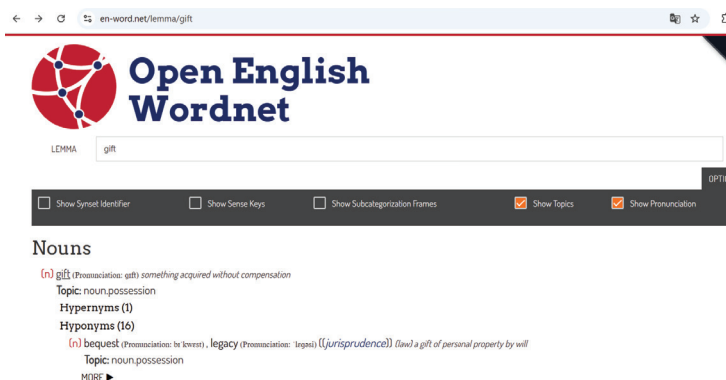
– tezaurus yondashuvi nafaqat alohida leksik birliklarning tavsifini, balki ular o'rtasidagi semantik munosabatlarning doimiy tavsifini ham nazarda tutadi. Ya'ni, tezaurus yondashuvi yordamida ta'riflardan birida "semantik maydon" deb nomlangan tushuncha shakllanadi. Tezaurus lug'at maqolasining tipik tarkibi bunga misol bo'la oladi: 1) bosh atama; 2) ta'rif; 3) sinonimlar; 4) holonimlar; 5) komponent, belgi, funksiya, aspekt, xususiyat munosabat nomlarini o'z ichiga olgan meronimlar; 6) giperonimlar; 7) giponimlar; 8) ekvivalent.

Ko'rinib turibdiki, terminning o'zi va uning ta'rifi va (ixtiyoriy ravishda) ruscha ekvivalenti tezaurus lug'at maqolasining yarmidan kamrog'ini tashkil qiladi. Asosiy e'tibor aynan ko'rib chiqilayotgan terminning muayyan terminosistema tarkibidagi boshqalar bilan semantik munosabatlarining turlariga qaratilgan.

– Ikkinchidan tezaurusga asoslangan yondashuv muayyan darajadagi iyerarxik ko'rinishidagi soha doirasidagi leksik-semantik strukturani ifodalash imkonini beradi. Ushbu aspektda leksik birliklar o'rtasidagi masofani o'lchash imkoniyatini yaratadi.

– Uchinchidan tezaurus orqali amaliy maqsadlarga yo'naltirilgan istalgan soha doirasida leksik-semantik strukturani amaliy jihatdan samarali modellashtirish imkonini beradi (ayniqsa, axborot qidiruv tizimida)

Ingliz tili WordNeti quyidagicha qidiruv interfeysiga ega [<https://en-word.net/>]:



1-rasm. Ingliz tili WordNeti

Ushbu lug'atda giponim va giponimik munosabatlarni ingliz tilida *gift* so'z orqali tahlil qilganimizda o'zbek tili bilan bog'liq ravishda tushunarli bo'lishi uchun uning o'zbek tilidagi misolini keltirishga harakat qilamiz.

1-jadval:

Sovg'a	– kompensatsiyasiz olingan mukofot	
Giperonim (1)		
Giponim	hadya (yuridik)	bir shaxsning o'z mulkini 2-shaxsga bepul berishi haqidagi shartnoma.
	xayr-ehson	Umumfoydali maqsadlarda qilingan hadya xayr-ehson hisoblanadi
	meros	Marhumdan vorislarga qolgan mulk
	o'lja	To'lovsiz olingan narsa
	Mahr	Teng huquqli nikoh (zavoj) tuzilgan paytda xotiniga ajratib beradigan mulk
	tuhfa	Kimsaning ko'nglini olish yoki xizmatlarini taqdirlash uchun berilgan narsa; tortiq, sovg'a
	tortiq	Kimsaga tuhfa qilingan narsa; sovg'a, armug'on
	armug'on	Badiiy. Tuhfa, sovg'a

Ideografik lug'at sifatida tezaurus konseptlarga birikar ekan, bu borada N.Lukashevich [Lukashevich va boshq., 2018] tomonidan olib borilgan ishlariga murojaat etamiz. Unga ko'ra RuTezda munosabatlar quyidagilar bilan semantik munosabatga kirishadi:

- jismoniy obyektlar (qavat - bino) o'rtasida;
- mintaqalararo (Yevropa - Yevroosiyo);
- moddalar o'rtasida;
- to'dalar (batalon - rota) o'rtasida;
- matn qismlari orasida (bo'g'in -she'r) [Guarino, 2009];
- jarayonlar o'rtasida (ishlab chiqarish sikli -sanoat ishlab chiqarishi).

Shuningdek, N.Guariono biri boshqasiga bog'liq bo'lgan ichki subyektlar o'rtasidagi aloqalar uchun ham qism-butun munosabatlari o'rnatilishi mumkinligini e'tirof etadi:

- subyektning xususiyatlari (ko'chirish - kema);
- jarayondagi rol (investor - investitsiya);
- faoliyat sohasining ishtirokchisining faoliyat sohasi (sanoat korxonasi - sanoat).

Bundan tashqari asosiy munosabatlardan biri RuTez lug'atida ontologik munosabat bog'liqligini ko'rsatuvchi bog'liqlik bir tushunchaning ikkinchi tushunchada mavjudligini ko'rsatadi. Masalan, yuqorida qayd qilingan manba mualliflari bunday munosabatga kirishuvchi misol sifatida *daraxt va o'rmon* tushunchalarni keltiradi. Bu yerda *daraxt* leksemi konsept o'rnida o'rmon mavjud bo'lishini taqozo etadi. *Daraxt - o'rmon*, bu yerda *o'rmon* mavjud bo'lishini taqozo etuvchi tobe tushuncha sanaladi.

Tezaurus termini keng ma'noda turli tizimli tillar uchun semantik munosabatlarni nafaqat leksik, balki sohalar kesimida o'rganish uchun zarur sanaladi.

YUNESKO tezaurusini ko'p tilli tezaurus sifatida baholash mumkin. Ushbu tizim ingliz, fransuz, ispan va rus tillaridagi hujjatlarni tasniflash va tahlil qilish tizimi uchun mo'ljallangan. Uning tarkibiy qismi sifatida rasmiylashtirilgan atamalar ro'yxati, hujjatlar va nashrlarni mavzuli tahlil qilish uchun ishlatiladi. Mazkur tizim o'z navbatida mikrotezauruslarga bo'lingan bo'lib, yettita asosiy fan soha doirasida sohalar bilan tanishish imkonini beradi va ularga quyidagilar kiradi:

- ta'lim;
- tabiiy fanlar;
- madaniyat;
- ijtimoiy va gumanitar fanlar;
- axborot va kommunikatsiya;
- siyosat, huquq va iqtisodiyot;
- mamlakatlar va mamlakatlar guruhlar;

Ko'rsatilgan manbada tezaurusning asosiy maqsadi sifatida quyidagilar qayd etilgan:

- obyektlarni tavsiflashda izchillikni ta'minlash;
- onlayn katalogdan ma'lumotlarni qidirishni osonlashtirish.

Tezaurus strukturasi deskriptorlardan va ular o'rtasidagi munosabatlardan iborat. Asosiy atamani tavsiflovchi deskriptorlar, leksik birliklar. (alohida so'z va so'z birikmalari bo'lishi mumkin.) Deskriptorlar alifbo bo'yicha iyerarxik tartibda berilib, ular iyerarxiya munosabat asosida loyihalashtirilgan.

Rus tili uchun RuThes tezaurusini alohida ta'kidlash o'rinli. U turli axborot-qidiruv tizimlarda, matnlarni umumlashtirish,

klasterlash, qisqartish va NLPning turli yo'nalishlarida ijobiy natijalar olishda muhim rol o'ynagan.

Ayniqsa, konseptlarni ifodalash uchun RuTez modelidan foydalanib, bir nechta domenga xos tezauruslar NLP va domen uchun yaratilgan axborot-qidiruv ilovalari yaratilgan, jumladan, sotsiopolitik tezaurus [Lukachevich va Dobrov, 2015], Tabiiy fanlar ontologiyasi va texnologiya [Dobrov va Lukachevich, 2006], Bank tezaurusi [Nokel va Lukachevich, 2016] va boshqalar. Hozirda, RuThes tushunchalari yaratish uchun asos yaratadi. Tatar ijtimoiy-siyosiy tezaurusi [Galieva et al., 2017] N.Lukashevich qayd etishicha, oraliq yondashuv sifatida tadqiqotchilar yangi til uchun ikki bosqichli so'z tarmog'ini yaratishi taklif qilingan: birinchi navbatda Princeton WordNet (WordNet yadrosi deb ataladigan) asosiy tushunchalarining munosabatlarini tarjima qilish va uzatish, so'ngra lug'atlar va matn korpuslari asosida iyerarxiyalarni qo'l mehnati orqali to'ldirish. Ushbu yondashuv DanNet kabi resurslarni yaratishda qo'llanilgan.

Birinchidan, RuTezda so'z turkumlari bo'yicha leksik tarmoqlarga bo'linish kuzatilmaydi. Nutqning har qanday qismi bir xil RuTez bilan bog'lanishi mumkin.

RuTezda munosabatlar faqat tushunchalar o'rtasida o'rnatilgan bo'lib, asosiy sinf-kichik sinf munosabati WordNetdagi giponim-giperonim munosabatiga deyarli mos keladi. Shuningdek, RuTez qism-butun munosabatga ega, ammo WordNetdan farqli o'laroq, bir nechta muqobil butunga emas, balki butunlikka ishora qiladi.

Demak, har bir leksik qidiruv tizim uchun tezaurus modeli va uning semantik tarkibi muhim ahamiyat kasb etib, ayniqsa, axborot qidiruv tizimi uchun muhim lingvistik resurs vazifasini bajaradi [Agostini, 2021].

Xulosa

Tezauruslar til va semantik munosabatlarni o'rganishda muhim vositalar bo'lib, axborot-qidiruv tizimlarida ma'lumotlarni izchil tasniflash va qidirishda katta ahamiyatga ega. YUNESKO tezaurusi ko'p tilli tizim sifatida turli tillarda hujjatlarni tahlil qilish uchun ishlatiladi, RuTez esa axborot-qidiruv tizimlarida samarali ishlatilgan. RuTezda tushunchalar o'rtasidagi munosabatlar giponim-giperonim va qism-butun aloqalarida ifodalanadi. Bu tizimlar, leksik va semantik ma'lumotlarni tizimlashtirib, axborot qidirish va tushunchalar o'rtasidagi aloqalarni aniq ifodalashda muhim rol o'ynaydi. Umuman, tezauruslar lingvistik resurslar sifatida axborot texnologiyalari va sun'iy intellekt sohalarida ham keng qo'llaniladi.

Foydalanilgan adabiyotlar

- Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the Uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
- George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller Introduction to WordNet: An On-line Lexical Database <https://wordnetcode.princeton.edu/5papers.pdf>
http://db4.sbras.ru/elbib/data/show_page.phtml?20+1530
<https://en-word.net/>
- Natalia Loukachevitch, German Lashevich and Boris Dobrov (2018) Comparing Two Thesaurus Representations for Russian, Proceedings of the 9th Global WordNet Conference (GWC 2018), pp 35–44, Singapore, ISBN 978-981-11-7087-4
- Natalia Loukachevitch, German Lashevich and Boris Dobrov (2018) Comparing Two Thesaurus Representations for Russian, Proceedings of the 9th Global WordNet Conference (GWC 2018), pp 35–44, Singapore, ISBN 978-981-11-7087-4
- Nicola Guarino. 2009. The Ontological Level: Revisiting 30 Years of Knowledge Representation. In Conceptual Modeling: Foundations and Applications: Essays in Honor of John Mylopoulos. Lecture Notes in Computer Science, 5600, Berlin and Heidelberg, Germany: Springer-Verlag: 52–67.
- Алексеев А.А., Добров Б. Лукашевич Н. Лингвистическая онтология – тезаурус РуТез. OSTIS-2013. – С. 153-154.
- Журавлев А. Тезаурусный подход как инструмент моделирования лексикосемантической структуры терминополья Известия Самарского научного центра Российской академии наук. Социальные, гуманитарные, медико-биологические науки, т. 23, № 80, 2021 Izvestiya of the Samara Science Centre of the Russian Academy of Sciences. Social, Humanitarian, Biomedical Sciences, Vol. 23, no. 80, 2021
- Осокина С. Основания лингвистической теории тезауруса: автореф. док фил.наук, Челябинск, 2015. – С. 20.

RESEARCH ON THESAURI IN LEXICOGRAPHIC SOURCES

Gulruxsor Xoliyorova¹

Abstract. This article discusses thesauruses, their importance in creating dictionaries, and principles of classification. The advantages of thesaurus dictionaries are explained with the help of examples.

Thesaurus dictionaries are databases created to analyze words and their relationships. They are collections of words that are close to each other. The main purpose of these dictionaries is to find other words that have the same meaning as a word and to find words that are similar to them. They can be used in texts or in various databases. Thesaurus dictionaries are widely used to help analyze texts and translate them. They display a list of similar or related words in a text. This helps to find the word you are looking for in the text. Thesaurus dictionaries are also used to classify texts and to organize taxa. Thesaurus dictionaries offer several useful features compared to other dictionaries. They help analyze the meanings of words, identify their relationships, explain words and their relationships, and classify texts. This is very useful in our work with texts and various databases. Thesaurus dictionaries, although at first glance they seem to be the same as dictionaries, have distinctive features.

Keywords: *thesaurus, dictionary, semantic connection, WordNet.*

References

- Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the Uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
- Alekseev A.A., Dobrov B. Lukashovich N. Lingvisticheskaya ontologiya – tezaurus RuTez. OSTIS-2013. – C. 153-154.
- George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller Introduction to WordNet: An On-line Lexical Database <https://wordnetcode.princeton.edu/5papers.pdf>

¹Xoliyorova Gulruxsor G'ayratovna - teacher at the Department of Computer Linguistics and Applied Linguistics, National University of Uzbekistan.
E-mail: guliholiyorova@gmail.com

- Juravlev A. Tezaurusnyy podkhod kak instrument modelirovaniya leksikosemanticheskoy struktury terminopolya Izvestiya Samarskogo nauchnogo sentra Rossiyskoy akademii nauk. Sotsialnye, gumanitarnye, mediko-biologicheskie nauki, t. 23, № 80, 2021 Izvestiya of the Samara Science Centre of the Russian Academy of Sciences. Social, Humanitarian, Biomedical Sciences, Vol. 23, no. 80, 2021
- Natalia Loukachevitch, German Lashevich and Boris Dobrov (2018) Comparing Two Thesaurus Representations for Russian, Proceedings of the 9th Global WordNet Conference (GWC 2018), pp 35–44, Singapore, ISBN 978-981-11-7087-4
- Natalia Loukachevitch, German Lashevich and Boris Dobrov (2018) Comparing Two Thesaurus Representations for Russian, Proceedings of the 9th Global WordNet Conference (GWC 2018), pp 35–44, Singapore, ISBN 978-981-11-7087-4
- Nicola Guarino. 2009. The Ontological Level: Revisiting 30 Years of Knowledge Representation. In *Conceptual Modeling: Foundations and Applications: Essays in Honor of John Mylopoulos*. Lecture Notes in Computer Science, 5600, Berlin and Heidelberg, Germany: Springer-Verlag: 52–67.
- Osokina S. Osnovaniya lingvisticheskoy teorii tezaurusa: avtoref. dok fil.nauk, Chelyabinsk, 2015. – С. 20

Jurnal 2017-yil 26-oktyabrda O'zbekiston Respublikasi Matbuot va axborot agentligi tomonidan 0936-raqam bilan ro'yxatdan o'tgan.

Jurnal O'zbekiston Respublikasi Oliy Attestatsiya Komissiyasi tomonidan filologiya fanlari bo'yicha falsafa doktori (PhD) va fan doktori (DSc) dissertatsiyalari asosiy ilmiy natijalari chop etilishi lozim bo'lgan ro'yxatga kiritilgan (30.10.2021. № 308/6).

Tahririyatga kelgan maqolalar mualliflarga qaytarilmaydi.

Manzil: Toshkent shahri, Yakkasaroy tumani, Yusuf Xos
Hojib ko'chasi 103-uy.
Telefonlar: +99871 281-45-11, +99871 281-41-93.
Website: compling.tsuull.uz
E-mail: kompling@navoiy-uni.uz

Bosishga **.**.****-yilda ruxsat etildi.
Bichimi 70x100 1/16, Ofset bosma. "Cambria" garniturasida.
Shartli b.t. 7,51. Nashr b.t. 7,62.

"O'zbekiston: til va madaniyat" jurnali tahririyatida tayyorlandi va sahifalandi.
"YASHNOBOD NASHR" bosmaxonasida chop etildi.
Adadi 300 nusxa. Buyurtma №2.
Bosmaxona manzili: Toshkent shahar Yashnobod tumani,
58-a harbiy shaharcha.