

AGGLYUTINATIV TIL UCHUN ELEKTRON TEZAVURUS YARATISH: O‘ZBEK TILI MISOLIDA ALGORITM VA ARHITEKTURA TAHLILI

Vohidova Barchinoy Baxtiyor qizi

Kompyuter lingvistikasi yo‘nalishi talabasi

barchinoyvohidova39@gmail.com

Ilmiy rahbar: Zilola Xusainova

ToshDO‘TAU

Annotatsiya. Ushbu maqolada agglyutinativ tillar, xususan o‘zbek tili uchun elektron tezavrus yaratishning arxitekturasini, algoritmi va leksikografik tuzilishi birinchi bor tizimli tahlil qilinadi. Mavjud tezavrus modellarining o‘zbek tiliga muvofiqlik darajasi ko‘rib chiqiladi, sinset-asosli va morfologik-asosli yondashuvlarning afzalliklari va kamchiliklari solishtirilib, o‘zbek tili uchun gibridd arxitektura taklif etiladi. Jarayonning to‘rt bosqichi – ma’lumot yig‘ish, leksikografik tahlil, semantik modellash va tezavrus chiqishi – BPMN diagramma asosida ko‘rsatiladi. Semantik munosabatlar modeli o‘zbek tili misollarida tavsiflanadi.

Kalit so‘zlar: elektron tezavrus; agglyutinativ til; WordNet; sinset; leksikografik arxitektura; semantik munosabatlar; morfologik tahlil; o‘zbek tili; BPMN algoritmi.

Kirish

Elektron tezavrus yaratish – leksikografiyaning eng murakkab va resurs talab qiladigan yo‘nalishlaridan biri. Ingliz tili, nemis tili, turk tili va boshqa bir qator tillar uchun bunday tizimlar allaqachon ishlab chiqilgan bo‘lsa-da [3, 4], agglyutinativ tillarning o‘ziga xos xususiyatlari tezavrus arxitekturasiga qo‘shimcha talablar qo‘yadi. O‘zbek tili – 35 milliondan ortiq kishi so‘zlaydigan agglyutinativ til – bu muammoning yaqqol namunasidir [5, 9].

Mavjud tadqiqotlar tezavrusning zarurligi haqida bir ovozdan gapiradi [6, 9]. Biroq ularda hal etilmagan muhim savol qolayapti: bunday tezavrusni qanday qilib yaratish kerak? Qaysi arxitektura o‘zbek tiliga mos keladi? Algoritmi qanday

bosqichlardan iborat bo'lishi kerak? Semantik munosabatlar qanday modellanadi? Aynan shu savollarga javob berish ushbu maqolaning maqsadidir.

Maqolaning ilmiy yangiligi uchta jihatdan ifodalanadi. Birinchidan, agglyutinatив tillar uchun tezaurus yaratish algoritmi birinchi bor BPMN formatida ko'rsatiladi. Ikkinchidan, o'zbek tiliga xos leksikografik arxitektura – gibril sinset-morfologik model – taklif etiladi. Uchinchidan, semantik munosabatlar tizimi o'zbek tili misollarida batafsil tavsiflanadi.

Adabiyotlar sharhi

Tezaurus nazariyasi Peter Mark Roget (1852) ning ideografik lug'atidan boshlanadi [7]. Uning modeli so'zlarni alifbo tartibida emas, mavzular bo'yicha guruhlaydi. Zamonaviy elektron tezauruslar esa bu g'oyani juda uzoqqa olib ketdi: Princeton University tomonidan ishlab chiqilgan WordNet (Fellbaum, 1998) sinset-asosli model bilan tezaurus leksikografiyasini tubdan o'zgartirdi [3]. WordNet da har bir tushuncha sinonimlar to'plami (sinset) sifatida ifodalanib, boshqa sinsetlar bilan semantik munosabatlar orqali bog'lanadi.

Vossen (1998) ning EuroWordNet loyihasi bu yondashuvni ko'p tilga kengaytirdi [4]. Yevropa tillarida WordNet bazalari yaratish tajribasi shuni ko'rsatdiki, har bir til o'ziga xos leksik-semantik tuzilishga ega va universal model to'liq ishlamaydi. Ayniqsa bu masala agglyutinatив tillar uchun keskinlashadi.

O'zbek tili uchun UzWordNet (Agostini va boshq., 2021) birinchi katta qadam bo'ldi [5]. “Expand” metodida yaratilgan bu baza taxminan 12 000 sinsetni qamraydi. Biroq unda ikki jiddiy muammo bor: morfologik shakl variativligi hisobga olinmagan va sinonimiya munosabatlari o'zbek tiliga xos tarzda modellanmagan. Pulatova (2025) sifat turkumi bazasini yaratishda bu muammolarni qisman hal etishga harakat qilgan [6], biroq butun leksik-semantik tizim uchun yechim hali taklif etilmagan. 1-jadval asosiy adabiyotlarni qiyosiy tarzda ko'rsatadi:

1-jadval. Tezaurus tadqiqotlari va ularning o'zbek tiliga muvofiqligi

Muallif(lar)	Yil	Asosiy hissa	Chegara	Manba
Fellbaum	1998	WordNet – sinset asosli leksik-semantik baza	Agglyutinatив tillar uchun cheklangan	[3]
Vossen	1998	EuroWordNet: ko‘p tilli kengaytma	Yevropa tillari bilan cheklangan	[4]
Agostini va boshq.	2021	UzWordNet: o‘zbek tili uchun birinchi leksik-semantik baza	Qamrov cheklangan, morfologiya hisobga olinmagan	[5]
Pulatova	2025	UzWordNet uchun sifat turkumi bazasi	Faqat sifat turkumi bilan cheklangan	[6]
Abduraxmonova, Xoliyorova	2024	Tezaurus yaratishda standart me‘yorlar tahlili	Algoritm va arxitektura taklif etilmagan	[9]

Jadvaldan ko‘rinib turibdiki, mavjud barcha ishlar yoki universal modellarga asoslangan, yoki o‘zbek tili uchun cheklangan qamrovda qolgan. Agglyutinatив til uchun maxsus yaratilgan, to‘liq algoritmgaga ega tezaurus modeli hali yo‘q – tadqiqot bo‘shlig‘i shu yerda.

Tezaurus turlari va o‘zbek tiliga muvofiqligi

Leksikografiya nazariyasida tezaurusning bir necha asosiy turi ajratiladi. Ularning o‘zbek tiliga muvofiqlik darajasi bir xil emas – bu masalani alohida ko‘rib chiqish zarur [2, 7]. 2-jadvalda asosiy tezaurus turlari taqqoslab ko‘rsatilgan:

2-jadval. Tezaurus turlari va o‘zbek tiliga muvofiqlik darajasi

Turi	Tamoyili	Misol	O‘zbek tiliga muvofiq	Manba
Ideografik	Mavzuviy-semantik	Roget’s Thesaurus (1852)	O‘rtacha – navigatsiya qiyin	[2, 7]

Sinset- asosli	Sinonimlar to‘plami	Princeton WordNet	Cheklangan – sinonimiya kam	[3, 5]
Ontologik	Ierarxik tushunchalar	FrameNet, OWL bazalari	Kuchli – semantik boylik yuqori	[4, 8]
Morfologik	O‘zak + qo‘shimcha ierarxiyasi	Turk tili WordNet	Eng mos – agglyutinatativ tillar uchun	[6, 9]

Jadval ko‘rsatganidek, o‘zbek tili uchun eng mos yondashuv morfologik tezaurus modelidir – agglyutinatativ tillar uchun maxsus ishlab chiqilgan. Biroq mavjud tajriba asosan turk tili bilan cheklangan [6, 9]. Shu sababli, ushbu maqolada gibrid model taklif etiladi: sinset-asosli tuzilish (WordNet dan) + morfologik ierarxiya + ontologik munosabatlar (FrameNet dan).

Elektron tezaurusning leksikografik arxitekturas

A. Makrotuzilish

Makrotuzilish – tezaurusdagi barcha lug‘at birliklarining umumiy tartiblanish tamoyili. Taklif etilgan gibrid arxitektura makrotuzilish ikki qatlamdan iborat: (1) mavzuviy-semantik qatlam – tushunchalar semantik soha bo‘yicha guruhlangan; (2) alifbo-indeks qatlami – har qanday so‘z shaklidan tegishli mavzuviy guruhga o‘tish imkonini beradi [2, 7]. Ikkala qatlamning birlashuvi qog‘oz tezaurusiga nisbatan asosiy ustunlik: foydalanuvchi ham mavzu bo‘yicha, ham shakldan qidirishni amalga oshira oladi.

B. Mikrotuzilish

Mikrotuzilish – har bir alohida lug‘at maqolasining ichki tuzilishi. Taklif etilgan modelda har bir leksik birlik uchun quyidagi maydonlar majburiy: (1) lemma – standartlashtirilgan asosiy shakl; (2) morfologik ma’lumot – turkum, o‘zak, qo‘shimcha modeli; (3) semantik munosabatlar – sinonimlar, antonimlar, giperonimlar, giponimlar; (4) ta’rif – qisqa izoh; (5) kontekstual misol; (6)

frekventlik – korpusdagi uchrash chastotasi [3, 8]. Oxirgi ikki maydon an'anaviy tezauruslarda bo'lmaydi – bu elektron formatning o'ziga xos qo'shimchasi.

S. Agglyutativ morfologiya uchun maxsus qatlam

O'zbek tiliga xos muammo shundaki, bitta o'zakdan yuzlab shakllar hosil bo'lishi mumkin. “Yoz-” o'zagidan: yozmoq, yozilmoq, yozdirmoq, yozuvchi, yozuv, yozma va boshqalar. Bularning har birini alohida sinset sifatida ko'rsatish tezaurusni haddan kattalashtiradi; barchasini bir sinsetga qo'shish esa semantik farqlarni yo'qotadi [6, 9]. Yechim: lemma-asosli ierarxiya – har bir o'zak bitta asosiy sinsetga ega, undan hosil bo'lgan shakllar esa ikkilamchi sinsetlar sifatida bog'liq holda ko'rsatiladi.

Semantik munosabatlar modeli

Elektron tezaurusning asosiy qimmati semantik munosabatlar tizimida. Oddiy izohli lug'atdan farqi ham aynan shu: tezaurus so'zlarning ma'nosini emas, so'zlar orasidagi munosabatlarni ko'rsatadi [3, 7]. 3-jadvalda o'zbek tili uchun taklif etilgan semantik munosabatlar tizimi misollar bilan keltirilgan:

3-jadval. O'zbek tili tezaurusi uchun semantik munosabatlar tizimi

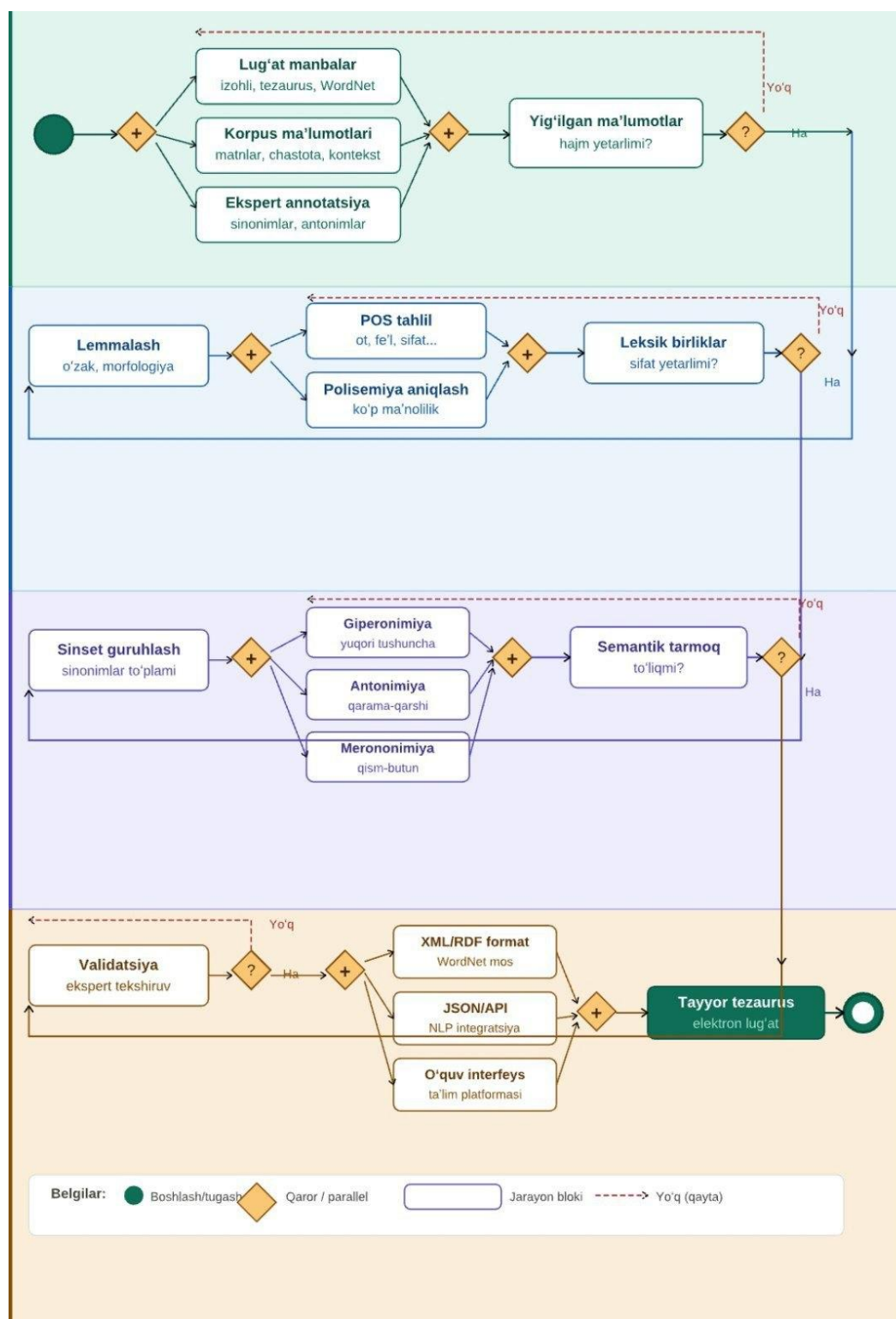
Munosabat	Belgilash	O'zbek misoli	Manba
Sinonimiya	$A \approx B$ (bir referent)	baxt $\approx$ saodat $\approx$ farovonlik	[3, 7]
Antonimiya	$A \leftrightarrow B$ (qarama-qarshi)	baxt $\leftrightarrow$ baxtsizlik	[3, 7]
Giperonimiya	$A \rightarrow B$ (A yuqori tushuncha)	his-tuyg'u $\rightarrow$ baxt	[3, 5]
Giponimiya	$A \leftarrow B$ (A quyi tushuncha)	baxt $\leftarrow$ oilaviy baxt	[3, 5]
Merononimiya	$A \subset B$ (qism-butun)	qo'l $\subset$ tana	[3, 8]
Polisemiya	$A \rightarrow \{m1, m2, m3\}$	ot: hayvon / ot: transport / ot: grammatika	[4, 10]



Jadvalda keltirilgan oltita munosabat tizimi o'zbek tili leksikasining semantik boyligini to'liq qamrab olishga imkon beradi. Ayniqsa polisemiya munosabati o'zbek tili uchun alohida ahamiyatli: bir shaklning bir necha ma'nosi bo'lgan holatlarda tezaurus bu ma'nolarni alohida sinsetlarga bo'lib, ular orasidagi chegara chiziqlarini aniq ko'rsatadi [4, 10]. Hozircha UzWordNet da polisemiya to'liq modellanmagan – bu taklif etilgan modelning eng muhim farqlangan tomoni.

Tezaurus yaratish algoritmi: BPMN tahlili

Elektron tezaurus yaratish iterativ jarayon bo'lib, to'rt asosiy bosqichdan iborat. 1-rasm ushbu jarayonning to'liq BPMN diagrammasini ko'rsatadi.



1-rasm. O'zbek tili elektron tezaurusini yaratishning BPMN diagrammasi

A. 1-bosqich: Ma'lumot yig'ish

Birinchi bosqichda uchta parallel manba bir vaqtda ishlatiladi: lugʻat manbalar, korpus maʼlumotlari va ekspert annotatsiya [3, 9]. Uch manbaning birligi muhim: faqat lugʻatga tayanish eskirgan maʼlumot beradi, faqat korpusga tayanish esa notoʻgʻri sinonimlar kiritishi mumkin. Yigʻilgan maʼlumotlar hajmi va sifati yetarli deb topilsa, ikkinchi bosqichga oʻtiladi; aks holda manbalar kengaytiriladi.

B. 2-bosqich: Leksikografik tahlil

Ikkinchi bosqichda yigʻilgan materiallar leksikografik jihatdan qayta ishlanadi. Parallel ravishda POS tahlil (grammatik turkumni aniqlash) va polisemiya aniqlash (koʻp maʼnoli soʻzlarni ajratish) amalga oshiriladi [4, 8]. Oʻzbek tilida bu bosqich ayniqsa murakkab, chunki grammatik turkumlarning chegarasi baʼzan noaniq: “yozuv” ot ham, sifat ham boʻla oladi.

S. 3-bosqich: Semantik modellash

Uchinchi bosqich – eng koʻp mehnat talab qiladigan qism. Sinset guruhlashdan boshlanib, giperonimiya, antonimiya va merononimiya munosabatlari parallel belgilanadi [3, 8]. Semantik tarmoq toʻliqligi tekshirilgach oxirgi bosqichga oʻtiladi.

D. 4-bosqich: Tezaurus chiqishi

Toʻrtinchi bosqichda tayyor tezaurus uch parallel formatda chiqariladi: XML/RDF (xalqaro muvofiqlik), JSON/API (NLP integratsiyasi) va oʻquv interfeysi (taʼlim platformalari) [5, 6]. Validatsiya bosqichi muhim: ekspert tekshiruvidan oʻtmagan material chiqishga kirmaydi.

Muhokama va natijalar

Taklif etilgan gibril arxitektura mavjud yondashuvlardan bir necha jihatdan farq qiladi. Birinchidan, u oʻzbek tilining agglutinativ xususiyatini arxitekturaning markaziga qoʻyadi – morfologik qatlam ixtiyoriy qoʻshimcha emas, balki asosiy tuzilish elementi sifatida loyihalanadi [6, 9]. Bu turk tili WordNet tajribasidan

o'rgangan xulosa: flektiv tillar uchun yaratilgan model agglyutinatív tilga shunchaki ko'chirilsa, sinonimlar to'plami sun'iy kengayib, semantik aniqlik pasayadi.

Ikkinchidan, BPMN diagramma orqali ko'rsatilgan jarayon iterativ tarzda qurilgan: har bir bosqichda sifat tekshiruvi o'tkaziladi va kerak bo'lsa oldinga qaytish mexanizmi mavjud. O'zbek tilshunosligida bunday jarayon modeli ilgari taklif etilmagan [2, 7]. Uchinchidan, uch formatdagi chiqish tezaurusning bir vaqtda ilmiy, texnik va ta'limiy maqsadlarda ishlatilishini ta'minlaydi [5, 8] – hozircha UzWordNet faqat ilmiy maqsadda mavjud.

Tadqiqotning asosiy cheklovi – taklif etilgan arxitektura hali to'liq amalga oshirilmagan, ya'ni bu nazariy model. Kelajakda prototip yaratish, o'zbek tili leksikasining kamida bir semantik sohasini (masalan, his-tuyg'ular) ushbu model asosida kodlash va natijalarni UzWordNet bilan solishtirish keyingi bosqich bo'lishi kerak.

Xulosa

Ushbu maqolada agglyutinatív til uchun elektron tezaurus yaratishning arxitekturasi, algoritmi va leksikografik tuzilishi birinchi bor tizimli tahlil qilindi. Mavjud tezaurus modellari orasida morfologik-asosli yondashuv o'zbek tiliga eng mos keladi, biroq u hali amalga oshirilmagan [6, 9]. Taklif etilgan gibrid model o'zbek tilining leksik boyligini to'liq aks ettirish imkonini beradi.

BPMN diagrammada ko'rsatilgan to'rt bosqichli algoritm iterativ va sifatga yo'naltirilgan jarayonni ta'minlaydi. Semantik munosabatlar tizimi – sinonimiya, antonimiya, giperonimiya, giponimiya, merononimiya, polisemiya – o'zbek tili misollarida birinchi bor to'liq tavsiflanadi [3, 4, 10]. Kelajakdagi tadqiqotlar uchun: (1) taklif etilgan modelda prototip yaratish; (2) morfologik tahlillagich bilan integratsiya sinovlari; (3) ta'lim platformasi uchun interfeys moduli ishlab chiqish.

Foydalanilgan adabiyotlar ro'yxati:

1. Roget P. M. *Thesaurus of English Words and Phrases*. – London: Longman, 1852.
2. Zgusta L. *Manual of Lexicography*. – The Hague: Mouton, 1971. – 360 p.
3. Fellbaum C. (ed.). *WordNet: An Electronic Lexical Database*. – Cambridge, MA: MIT Press, 1998. – 423 p.
4. Vossen P. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. – Dordrecht: Kluwer, 1998. – 184 p.
5. Agostini A., Usmanov T., Khamdamov U., Abdurakhmonova N., Mamasaidov M. UZWORDNET: A Lexical-Semantic Database for the Uzbek Language // *Proceedings of the 11th Global WordNet Conference (GWC-2021)*. – P. 8–19.
6. Pulatova G. Designing a lexical database of adjective thesauri for the Uzbek WordNet // *International Journal of Literature and Languages*. – 2025. – Vol. 5(11). – P. 108–112.
7. Kurbonova F. *Kompyuter lug'atlari: tezaurus*. – Toshkent, 2014.
8. Oltmans E., Bunt H. The use of ontologies in the design of a semantic lexicon // *Proceedings of the 1st OntoLex Workshop*. – Las Palmas, 2002. – P. 1–8.
9. Abduraxmonova N., Xoliyorova G. O'zbek tili tezaurusini yaratishda standart me'yorlar // *MyScience.uz*. – 2024.
10. Lyons J. *Semantics* (Vols. 1–2). – Cambridge: Cambridge University Press, 1977.