

## UZWORDNET ONTOLOGIYASINING ARXITEKTURASI VA MA'LUMOTLAR MODELI

**Elov Botir Boltayevich,**  
texnika fanlari doktori (DSc), dotsent  
[elov@navoiy-uni.uz](mailto:elov@navoiy-uni.uz)  
ToshDO'TAU

**Axmedova Aynura Quanishbaevna,**  
1-kurs magistrant  
[aynuraaxmedova345@gmail.com](mailto:aynuraaxmedova345@gmail.com)  
ToshDO'TAU

**Annotatsiya.** Maqolada o'zbek tili uchun WordNet tipidagi leksik-semantik resurs (UzWordNet) arxitekturasi tahlil qilingan. O'zbek tilining agglyutinativ xususiyatlarini hisobga olib, moslashtirilgan ontologik model ishlab chiqilgan. Natijada sinset, lemma, so'z turkumi va gloss kabi ma'lumotlarni birlashtiruvchi, 3000 sinset, 10000 lemma va 10 turdagi semantik munosabatni qamab oluvchi ko'p qatlamli model taklif etilgan.

**Kalit so'zlar:** *UzWordNet, WordNet, leksik-semantik resurs, sinset, lemma, ma'lumotlar modeli, ontologiya, o'zbek tili, NLP*

**Abstract.** This article analyzes the architecture of a WordNet-type lexical-semantic resource for the Uzbek language, namely UzWordNet. Taking into account the agglutinative features of Uzbek, an adapted ontological model is developed. As a result, a multilayer model is proposed that integrates data such as synsets, lemmas, parts of speech, and glosses, and covers 3000 synsets, 10000 lemmas, and 14 types of semantic relations.

**Keywords:** *UzWordNet, WordNet, lexical-semantic resource, synset, lemma, data model, ontology, Uzbek language, NLP.*

**Kirish.** Raqamli til resurslari tabiiy tilni qayta ishlash tizimlarining muhim infratuzilmasi hisoblanadi. Morfologik tahlil, semantik qidiruv, mashina tarjimasini va matni tasniflash kabi vazifalarda so'zlar o'rtasidagi ma'no munosabatlari, konseptual yaqinlik va iyerarxik bog'lanishlarni ifodalovchi bilimlar bazasi talab



etiladi. WordNet konsepsiyasi ushbu ehtiyojga javob beruvchi model bo'lib, unda leksik birliklar ma'no markazi atrofida birlashgan sinsetlar sifatida tashkil qilinadi [9:39–41].

O'zbek tili uchun bunday resursni yaratish ikki jihatdan dolzarbdir. Birinchidan, o'zbek tili raqamli resurslar jihatidan boshqa tillarga nisbatan kamroq qamrovga ega. Ikkinchidan, tilning agglyutinativ morfologiyasi va derivatsion xususiyatlari tayyor WordNet modelini bevosita ko'chirishga imkon bermaydi. Agglyutinativ tillarda POS tagging va stemming masalalari maxsus yondashuvni talab qiladi [2:57–62]. Shuningdek, o'zbek tilida morfologik va sintaktik tahlil usullarining rivojlanishi ham bunday semantik resurslar uchun muhim texnologik asos yaratadi [4:439–444].

**Tadqiqot metodologiyasi.** UzWordNet arxitekturasi sinset markazli yondashuv asosida ishlab chiqilgan bo'lib, modelda sinset asosiy konseptual birlik, lemma sinsetga kiruvchi til birligi, sense esa lemma va sinset o'rtasidagi bog'lanish sifatida talqin qilinadi. Bu uch pog'onali yondashuv ko'p ma'noli so'zlar bilan ishlash va munosabatlarni aniq belgilash imkonini beradi. Tadqiqotda leksikografik manbalar: sinonim, izohli hamda o'quv lug'atlari metodik tayanch bo'lib xizmat qildi. Metodologik jarayon to'rt bosqichdan iborat bo'ldi: birinchi bosqichda leksik birliklar yig'ilib, normalizatsiya qilindi; ikkinchi bosqichda sinonimik qatorlar sinset nomzodlariga aylantirildi; uchinchi bosqichda sinsetlar semantik munosabatlar orqali bog'landi; to'rtinchi bosqichda ma'lumotlar bazasi formal struktura sifatida tashkil etildi. Tadqiqotda asosiy e'tibor arxitekturaviy qatlarga qaratilib, UzWordNetning Lexical Markup Framework (LMF) standartlari bilan moslashuvi va kelgusida NLP tizimlariga integratsiya qilish imkoniyatlari yoritildi.

**Adabiyotlar tahlili.** WordNet modeli dastlab ingliz tili uchun ishlab chiqilgan bo'lib, u leksik birliklarni sinonimik guruhlar – sinsetlar asosida tashkil etish va ularni semantik munosabatlar orqali bog'lashga tayanadi [6:39–41]. Keyingi



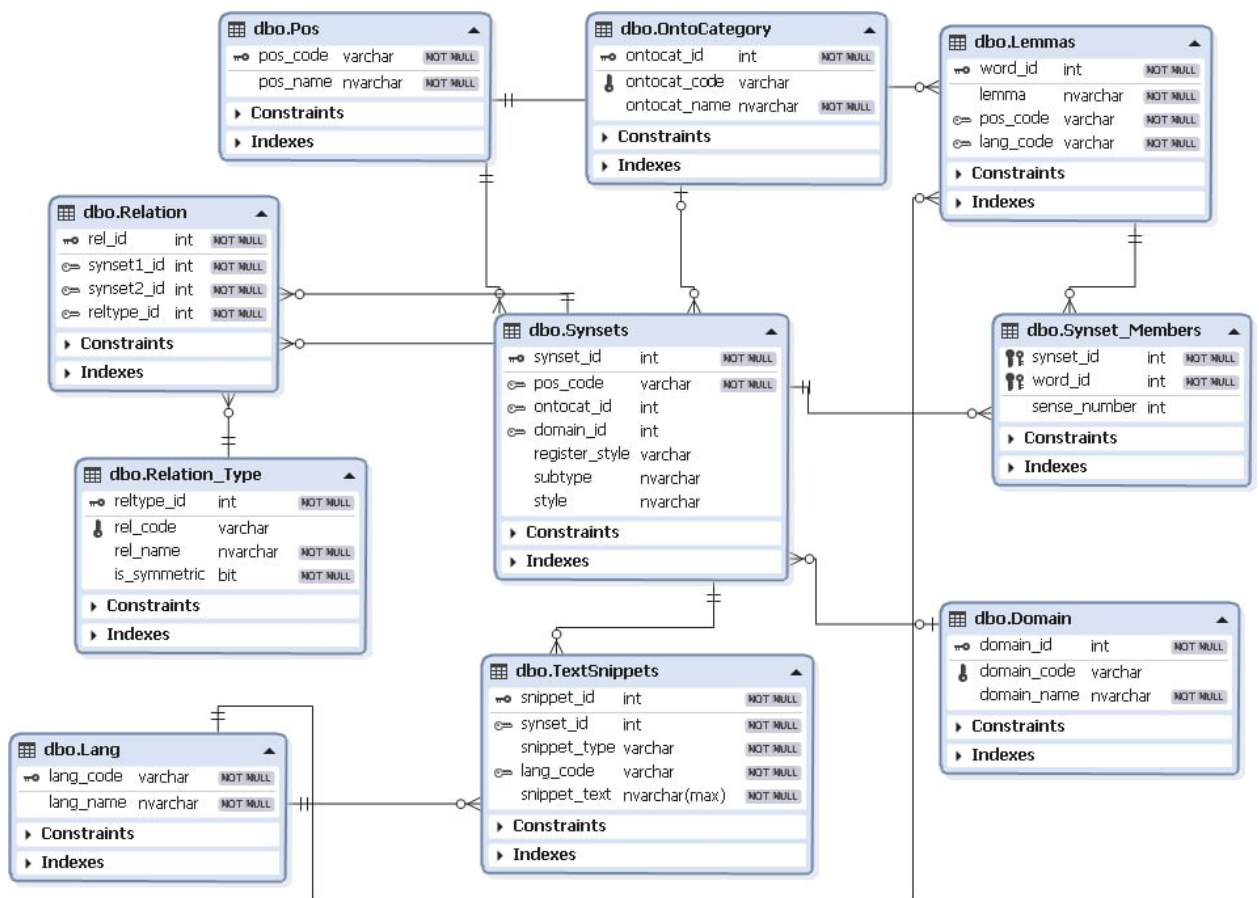
bosqichlarda EuroWordNet [10:73–89], Global WordNet [8] va Open Multilingual WordNet [4] kabi tashabbuslar ko‘p tilli moslik, interlingval indeks va resurslararo integratsiya muammolarini kun tartibiga olib chiqdi. Bu esa har bir milliy WordNet o‘z tilining semantik tizimini aks ettirishi bilan birga, xalqaro resurslar bilan bog‘lana olishini talab qiladi.

O‘zbek tili bo‘yicha mavjud ishlar ichida A. Agostini va boshqalar tomonidan taqdim etilgan UzWordNet loyihasi alohida ahamiyatga ega bo‘lib, o‘zbek tili uchun leksik-semantik ma’lumotlar bazasi yaratish masalasi ko‘rib chiqilgan [1:8–19]. Shuningdek, B. Elov va boshqalar tomonidan o‘zbek, uyg‘ur va turk tillari kabi agglyutinativ tillar uchun so‘z turkumlarini belgilash (POS tagging) va so‘z o‘zagini aniqlash muammolari o‘rganilgan [2:57–62]. Bu ishlar o‘zbek tili uchun semantik resurs yaratish mumkinligini ko‘rsatadi, biroq resursning ichki arxitekturasini izchil belgilash va tilga xos derivatsion qatlamni ajratish vazifasi saqlanib qolmoqda. Kam resursli tillar uchun WordNet yaratish bo‘yicha xalqaro tajribalar o‘zbek tilining tipologik xususiyatlariga moslashtirilishi lozim [3:565–580].

**Uzwordnet ma’lumotlar modelining asosiy komponentlari.** UzWordNet modeli bir nechta o‘zaro bog‘langan obyektlar tizimidan iborat. Sinsetlar jadvali konseptual yadro vazifasini bajaradi. Har bir sinset alohida identifikator, so‘z turkumi, gloss, domen belgisi va zarur hollarda ontologik toifa bilan tavsiflanadi. Lemma jadvali esa resursga kiritilgan kanonik leksik birliklarni saqlaydi. Lemma bilan sinset bevosita tenglashtirilmaydi, chunki bitta lemma bir nechta ma’noga ega bo‘lishi mumkin.

*Sinset\_Members* tipidagi bog‘lovchi qatlam lemma va sinset o‘rtasidagi sense darajasidagi munosabatni ifodalaydi. Aynan shu qatlam polisemiya holatlarida muhim ahamiyat kasb etadi: masalan, bir xil yoziladigan birlik turli gloss, domen yoki misol kontekstlariga ega bo‘lsa, u alohida sense sifatida boshqa sinsetga bog‘lanadi. Bu yondashuv ma’no aralashuvi xavfini kamaytiradi.

*Relation* va *Relation\_Type* obyektlari sinsetlararo munosabatlarni boshqaradi. Giperonimiya, giponimiya, antonimiya, meronimiya, xolonimiya, derivatsion bog‘lanish, domen teg, ko‘p tilli ekvivalentlik kabi munosabatlar alohida tur sifatida saqlanadi. Bunda munosabatning yo‘nalishi, simmetrik yoki nosimmetrikligi, iyerarxik qatlarga tegishliligi va qo‘llanish cheklovlari ham ko‘rsatilishi kerak.



**1-rasm. UzWordNet ma'lumotlar modeli.**

Qo‘shimcha qatlamlar sifatida *POS*, *Domain*, *OntoCategory*, *Lang* va *TextSnippets* obyektlari kiritiladi. POS so‘z turkumini, Domain tematik sohani, *OntoCategory* yuqori ontologik sinfni, *Lang* ko‘p tillilik imkoniyatini, *TextSnippets* esa gloss va misol gaplarni boshqarishga xizmat qiladi. Bunday tuzilma UzWordNetni faqat lug‘at emas, balki qayta ishlanishi mumkin bo‘lgan semantik ma'lumotlar bazasiga aylantiradi.

*1-jadval. UzWordNet ma'lumotlar modelining asosiy komponentlari*

| Komponent            | Vazifasi                                 | UzWordNetdagi ahamiyati                                |
|----------------------|--|--|
| <b>Sinset</b>        | Ma'no markazini ifodalaydi               | Sinonimik birliklarni konseptual guruhga birlashtiradi |
| <b>Lemma</b>         | Kanonik leksik birlikni saqlaydi         | Yozuv va morfologik variantlarni nazorat qiladi        |
| <b>Sense</b>         | Lemma va sinset bog'lanishini belgilaydi | Polisemiya holatlarini ajratadi                        |
| <b>Relation_Type</b> | Munosabat turini formal belgilaydi       | Iyerarxik va assotsiativ aloqalarni boshqaradi         |
| <b>Gloss</b>         | Sinset ta'rifini beradi                  | Ma'no chegarasini aniqlashtiradi                       |
| <b>Domain</b>        | Tematik sohani ko'rsatadi                | Terminologik va uslubiy qamrovni nazorat qiladi        |

**Natijalar.** Taklif etilgan arxitektura asosida UzWordNetning dastlabki versiyasi 3000 sinset va 10000 lemmadan iborat boshlang'ich qamrov sifatida tashkil etildi. Resursda 10 turdagi semantik munosabat minimal to'plam sifatida ajratildi. Bu to'plam WordNet amaliyotidagi asosiy munosabatlarni qamrab olishi bilan birga, o'zbek tiliga xos derivatsion va domen belgilarini ham alohida ko'rsatadi.

Sinsetlarning ichki tarkibi ma'no birligi tamoyili asosida tekshirildi. Sinonimik qatorlar avtomatik ravishda yakuniy sinset sifatida qabul qilinmadi; ular gloss, kontekst, so'z turkumi va semantik moslik asosida qayta ko'rib chiqildi. Bu yondashuv ayniqsa ko'p ma'noli leksik birliklar bilan ishlashda muhim bo'ldi.

Munosabatlar ikki yirik guruhda tashkil qilindi: iyerarxik va noiyerarxik munosabatlar. Iyerarxik munosabatlar konseptlar orasidagi yuqori-quyi bog'lanishni, noiyerarxik munosabatlar esa antonimik, derivatsion, sababiy, domen va assotsiativ aloqalarni ifodalaydi. Natijada UzWordNet chiziqli lug'at ro'yxatidan ko'ra murakkab semantik navigatsiya imkonini beruvchi graf modeliga yaqinlashadi.

Arxitektura JSON, CSV, RDF va kelgusida LMF kabi formatlarga eksport qilish imkonini nazarda tutadi. Bu esa resursni semantik qidiruv, lemmatizator,



sentiment tahlili, mashina tarjimasini va savol-javob tizimlariga ulash uchun qulay sharoit yaratadi.

2-jadval. UzWordNetning dastlabki qamrov ko'rsatkichlari

| Ko'rsatkich       | Dastlabki qiymat | Izoh   |
|-------------------|------------------|--|
| Sinsetlar         | 3000             | Boshlang'ich versiya uchun minimal konseptual qamrov                   |
| Lemmalar          | 10000            | Kanonik shaklda kiritilgan birliklar                                   |
| Munosabat turlari | 10               | WordNet asosiy munosabatlari va o'zbek tiliga xos qo'shimcha qatlamlar |
| Glosslar          | 10000            | Izohli lug'atlar va tahrirlangan ta'riflar asosida                     |
| Eksport           | CSV/JSON/RDF     | NLP tizimlariga ulash uchun rejalashtirilgan formatlar                 |

**Muhokama.** UzWordNet arxitekturasi asosiy ustunligi uning kengaytiriluvchanligidir. Resurs dastlab cheklangan qamrov bilan tuzilgan bo'lsa-da, modelning sinset markazli tuzilishi yangi lemma, gloss, misol gap, domen belgisi va munosabatlarni bosqichma-bosqich qo'shish imkonini beradi. Versiyalash tamoyili esa har bir kengaytirish bosqichini qayta tekshirish, xatolarni qayd etish va resurs sifatini nazorat qilishga yordam beradi.

Biroq modelda ayrim cheklovlar ham mavjud. Lug'atlarga tayanish leksikografik ishonchlilikni oshiradi, lekin zamonaviy korpuslarda uchraydigan yangi so'zlar, terminlar va kontekstual ma'nolarni to'liq qamrab olmasligi mumkin. Shuning uchun keyingi bosqichlarda korpus statistikasi, distributiv semantika va avtomatik sinset taklif etish mexanizmlarini qo'shish maqsadga muvofiq. Kam resursli tillarda avtomatik kengaytirish tajribasi bu jarayon uchun muhim metodik asos bo'lishi mumkin.

Yana bir muhim masala tashqi WordNetlar bilan moslashtirishdir. Interlingval indeks orqali bog'lanish UzWordNetni xalqaro resurslar bilan integratsiya qilishga xizmat qiladi. Ammo bunday moslashtirishda o'zbek tiliga xos milliy-madaniy konseptlar majburiy ravishda inglizcha konseptlarga tenglashtirib

yuborilmasligi kerak. Shu bois arxitektura avval ichki semantik izchillikni, keyin esa ko'p tilli moslikni ta'minlashga yo'naltirilgan.

**Xulosa.** Maqolada o'zbek tili uchun UzWordNet leksik-semantik resursining arxitekturaviy modeli ishlab chiqildi. Model sinset, lemma, sense, munosabat turi, gloss, domen va ontologik toifa kabi komponentlarni yagona struktura doirasida birlashtiradi. Bu yondashuv o'zbek tilining agglyutinativ va derivatsion xususiyatlarini hisobga olgan holda WordNet tipidagi resurs yaratishga metodik asos bo'lib xizmat qiladi.

Dastlabki natijalar UzWordNetning ko'p munosabatli graf sifatida shakllanishi mumkinligini ko'rsatdi. 3000 sinset va 10000 lemma asosidagi boshlang'ich qamrov kelgusida korpusga tayangan kengaytirish, avtomatik tekshiruv va ko'p tilli integratsiya uchun tayanch platforma vazifasini bajaradi.

Kelgusida resurs sifatini baholash uchun sinset izchilligi, iyerarxik barqarorlik, relatsion bog'langanlik, ma'no ajratish aniqligi va NLP vazifalaridagi amaliy samaradorlik kabi mezonlar asosida alohida eksperimental tadqiqot o'tkazish zarur.

### Foydalanilgan adabiyotlar ro'yxati

1. A. Agostini, T. Usmanov, U. Khamdamov, N. Abdurakhmonova, and M. Mamasaidov, “UZWORDNET: A lexical–semantic database for the Uzbek language,” in Proc. 11th Global Wordnet Conf. (GWC), Global Wordnet Association, 2021, pp. 8–19.
2. B. Elov, E. Adali, S. M. Khamroeva, O. X. Abdullayeva, Z. Y. Xusainova, and N. U. Xudayberganov, “The problem of POS tagging and stemming for agglutinative languages (Turkish, Uyghur, Uzbek languages),” in Proc. 8th Int. Conf. Computer Science and Engineering (UBMK), 2023, pp. 57–62.
3. C. Fellbaum, Ed., WordNet: An Electronic Lexical Database. Cambridge, MA, USA: MIT Press, 1998.



4. E. B. Boltayevich, A. O. Xolmo‘minovna, A. F. Sharipovna, and A. N. Ganiyevna, “Methods of morphological and syntactic analysis for the Uzbek language,” in Proc. 10th Int. Conf. Computer Science and Engineering (UBMK), 2025, pp. 439–444.
5. F. Bond and R. Foster, “Linking and extending an open multilingual Wordnet,” in Proc. ACL 2013 (System Demonstrations), 2013.
6. G. A. Miller, “WordNet: A lexical database for English,” Communications of the ACM, vol. 38, no. 11, pp. 39–41, 1995.
7. G. Francopoulo and C.-R. Huang, “Lexical Markup Framework (LMF): ISO standard for lexicons,” Lexicography, vol. 1, pp. 5–12, 2014.
8. Global WordNet Association, “Wordnets in the World.” [Online]. Available: <https://globalwordnet.org/resources/wordnets-in-the-world/>
9. P. Berangi, Z. Mousavi, H. Faili, and A. Shakery, “WordNet construction for under-resourced languages using personalized PageRank,” Digital Scholarship in the Humanities, vol. 36, no. 3, pp. 565–580, 2021.
10. P. Vossen, “Introduction to EuroWordNet,” Computers and the Humanities, vol. 32, pp. 73–89, 1998.