



UDK: 811.512

**Botir ELOV,**

*ToshDO'TAU dotsenti, PhD*

*E-mail: elov@navoiy-uni.uz*

**Aynura AXMEDOVA,**

*ToshDO'TAU magistranti*

*E-mail: aynuraaxmedova345@gmail.com*

*PhD Z.Xusainova taqrizi asosida*

### BABELNET: A MULTI-SOURCE SEMANTIC NETWORK

Annotation

This article analyzes the architecture and synset model of BabelNet as a multilingual lexical-semantic network built on multi-source integration. It examines the automatic linking mechanisms with resources such as WordNet, Wikipedia, and Wikidata, as well as the formation of concepts and named entities within the system. Special attention is given to the inclusion of Turkic languages – particularly Uzbek – into the BabelNet semantic graph and its significance for cross-lingual semantic representation. The findings demonstrate BabelNet's role as a global semantic infrastructure and its potential applications in multilingual NLP systems.

**Keywords:** BabelNet, semantic network, lexical-semantic model, synset, ontology, Wikipedia integration, WordNet, semantic relations, machine translation, WSD, NLP, conceptual modeling, multi-source integration.

### BABELNET: МНОГОРЕСУРСНАЯ СЕМАНТИЧЕСКАЯ СЕТЬ

Аннотация

В статье рассматриваются архитектура и модель синсетов BabelNet как многоязычной лексико-семантической сети, основанной на многоресурсной интеграции. Анализируются механизмы автоматической связи с такими ресурсами, как WordNet, Wikipedia и Wikidata, а также процессы формирования концептов и именованных сущностей. Особое внимание уделяется включению тюркских языков, в частности узбекского, в семантический граф BabelNet и значимости этого для межязыкового семантического представления. Результаты исследования демонстрируют роль BabelNet как глобальной семантической инфраструктуры и его потенциал в многоязычных NLP-системах.

**Ключевые слова:** BabelNet, семантическая сеть, лексико-семантическая модель, синсет, онтология, интеграция с Wikipedia, WordNet, семантические отношения, машинный перевод, WSD, NLP, концептуальное моделирование, многоресурсная интеграция.

### BABELNET: KO'P MANBALI SEMANTIK TARMOQ

Annotatsiya

Maqolada BabelNetning ko'p manbali integratsiyaga asoslangan ko'ptilli leksik-semantik tarmoq sifatidagi arxitekturasi va sinset modeli tahlil qilinadi. WordNet, Wikipedia, Wikidata kabi resurslar bilan avtomatik bog'lanish jarayoni hamda konseptlar va nomlangan obyektlarning shakllanish mexanizmi yoritiladi. Shuningdek, turkiy tillar, xususan o'zbek tilining BabelNet tarkibiga kiritilishi va uning semantik tizimdagi ahamiyati ko'rib chiqiladi. Tadqiqot natijalari BabelNetning global semantik infratuzilma sifatidagi roli va ko'p tilli NLP tizimlarida qo'llanish imkoniyatlarini ochib beradi.

**Kalit so'zlar:** BabelNet, semantik tarmoq, leksik-semantik model, sinset, ontologiya, Wikipedia integratsiyasi, WordNet, semantik relatsiyalar, mashina tarjimas, WSD, NLP, konseptual modellashirish, ko'p manbali integratsiya.

**Kirish.** Inson tili dunyoni idrok etish va uni tafakkurda aks ettirishning murakkab, ko'p qatlamli tizimi bo'lib, undagi so'zlarning shakli, ma'nosi, konseptual strukturalari hamda ular orasidagi semantik munosabatlar o'zaro uzviy bog'liq holda ishlaydi. Tabiiy tillarning shu darajadagi murakkabligini mashina uchun tushunarli ko'rinishga keltirish, til birliklarini formal modelga aylantirish hamda turli tillardagi konseptual farqlarni yagona semantik makonda ifodalash kompyuter lingvistikasi oldidagi asosiy ilmiy muammolardan biridir.

Hozirgi kunda sun'iy intellekt tizimlari, mashina tarjimas, semantik izlash vositalari, intellektual agentlar va katta til modellari (LLM)ning samaradorligi ko'p jihatdan tilning semantik tuzilmasi qay darajada to'g'ri modellashtirilganiga bog'liq. Shu sababli semantik tarmoqlar, leksik-semantik bazalar va ontologik modellar zamonaviy lingvistik va texnologik izlanishlarning asosiy yo'nalishiga aylangan.

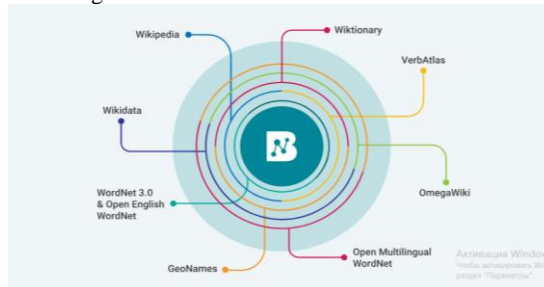
**Mavzuga oid adabiyotlar tahlili.** Tilning semantik tuzilishini formal ko'rinishda tasvirlashga qaratilgan dastlabki yirik loyihalardan biri WordNet bo'lib, u so'z ma'nolarini sinsetlar shaklida tashkil etib, ular orasidagi munosabatlar orqali ingliz tilining konseptual modelini yaratdi. Biroq global axborot oqimining kengayishi, ko'p tilli bilim manbalarining paydo bo'lishi, xususan Wikipedia va Wikidata kabi platformalarning ommalashuvi, shuningdek ko'p tilli sun'iy intellekt tizimlariga ehtiyojning ortishi WordNetning imkoniyatlarini cheklab qo'ydi. Endilikda so'zlarning ingliz tilidagi ma'nosini emas, balki yuzlab tillardagi ekvivalentlarini, qo'shimcha ravishda ensiklopedik, ontologik va multimodal bilimlarni ham yagona tizimda birlashtira oladigan resurs zarur edi.

**Tadqiqot metodologiyasi.** Tadqiqot BabelNetning tuzilmasi va ko'p manbali integratsiyasini o'rganishga qaratildi. Metodologiya ilmiy adabiyotlarni tahlil qilish,

sinsetlar va munosabatlarni strukturaviy o'rganish, turkiy tillar qamrovini statistik solishtirish hamda semantik bog'lanishlarni vizual tahlil qilish usullariga asoslandi.

**Tahlil va natijalar.** 2009-yilda Sapienza universiteti tadqiqotchisi Roberto Navigli tomonidan BabelNet loyihasi ishlab chiqildi [1]. BabelNet – WordNetning leksik-semantik

modeli, Wikipedia tarmoqlari, Wikidata ontologiyasi, OmegaWiki, Wiktionary, Open Multilingual WordNet va bir qator boshqa manbalardagi ma'lumotlarni avtomatik bog'lash asosida yaratilgan yirik ko'p tilli global ontologik tarmoqdir (1-rasmga qarang).



1-rasm. Babelnet 5.3 ga integratsiyalangan resurslar diagrammasi

BabelNetning ilk versiyasi 2013-yilda yaratilgan. BabelNetning versiyalari yildan-yilga sezilarli darajada kengayib, har bir yangi versiya semantik tarmoqning qamrovi, tillar soni, sinsetlar, leksemalar, ta'riflar va multimodal resurslar bo'yicha oldingi avloddan ancha ustun bo'lgan. 5.3 versiya (2023/12) hozirgacha eng katta va eng boy versiya hisoblanadi: 600 til, 1.7 milliarddan ortiq so'z ma'nolari, 22.8 million sinset va 7.3 million konsept bilan BabelNetni global miqyosdagi eng yirik ko'p tilli bilim grafigiga aylantirdi. Bu versiyada nomlangan obyektlar soni ham 15.5 million bo'lib,

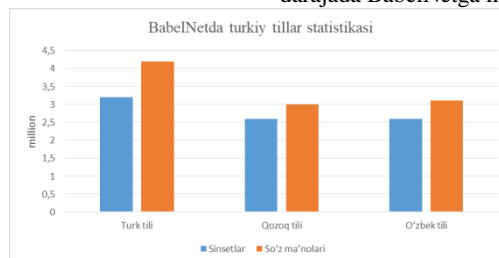
ensiklopedik obyektlar bilan bog'langan leksik birliklar tarmog'i nihoyatda kengaygan. Glosslar (izohlar) soni 159,6 millionga yetgani esa BabelNetni faqat sinsetlar emas, balki chuqur tavsiflangan konseptual modelga aylantiradi. Bundan oldingi versiyalar, masalan 5.2 (2022/11) va 5.1 (2022/07) ham sezilarli ko'rsatkichlarga ega bo'lsa-da, til qamrovi 520–500 til, sinsetlar esa 22.1–21.8 million atrofida bo'lgan. 5.0 versiyasi (2021/02) til qamrovi bo'yicha 500 tilda to'xtagan bo'lib, 20.3 million sinsetga ega edi. 4.0 versiyasi esa keskin sakrash davri bo'lib, 284 tildan darhol 500 tildan ortiq qamrovga o'tgan.

Ko'rsatkichlar	BabelNet 3.7	BabelNet 4.0	BabelNet 5.3
Tillar	271	284	600
So'z ma'nolari	745,859,932	808,974,108	1,706,278,218
Sinsetlar	13,801,844	15,780,364	22,892,310
Tushunchalar	6,066,396	6,113,467	7,327,078

BabelNetning markaziy tuzilmasini BabelSinset tashkil etadi. Ushbu sinsetlar o'z ichiga turli tillardagi lemmalar, glosslar, tarjima variantlari, Wikipedia sahifalariga havolalar, Wikidata identifikatorlari, multimodal obyektlar (masalan, rasm va boshqa vizual materiallar) hamda ko'plab semantik munosabatlarni jamlaydi. Bunday yondashuv natijasida so'zning faqat lug'aviy ma'nosi emas, balki uning konseptual, ontologik va real dunyo bilan bog'liq jihatlari ham izchil modellashiriladi.

Ontologiya tushunchasi esa BabelNetning ilmiy asosini belgilaydi. Ontologiya – ma'lum bir sohadagi tushunchalar, obyektlar va kategoriyalarni hamda ular orasidagi munosabatlarni formal ko'rinishda ifodalaydigan tizimdir. WordNet ham aslida leksik ontologiya sifatida shakllangan, unda sinsetlar konseptlar, ular orasidagi munosabatlar esa ontologik bog'lanish sifatida qaraladi.

BabelNetning muhim afzalligi shundaki, u faqat yevropa tillarigagina tayangan tizim emas; u dunyoning aksariyat yirik til oilalarini o'z ichiga oladi. Turkiy tillar ham bu jarayonda alohida ahamiyatga ega. Turkiy tillar morfologik jihatdan boy, agglutinatív tuzilishga ega va ko'plab so'zlarning polisemiya yuqori bo'lgani sabab, ularning semantik xaritasini yaratish oson jarayon emas. Turkiy tillar orasida turk tili eng katta qamrovga ega bo'lgan til sifatida ajralib turadi. Buning asosiy sababi – turk tilidagi Vikipediyaning nihoyatda katta hajmda bo'lishi. Turk tilining BabelNetdagi ko'lami 3 milliondan ortiq sinset va 4 milliondan ortiq so'z ma'nosini o'z ichiga oladi. Qozoq va o'zbek tillari esa hajm bo'yicha deyarli bir xil bo'lib, har ikkisi uchun 2 milliondan ortiq sinset va 3 milliondan ziyod so'z ma'nolari mavjud. Shuningdek, ozarbayjon, qirg'iz, uyg'ur, tatar, boshqird, turkman va hatto resurslari kam bo'lgan ayrim kichik turkiy tillar ham muayyan darajada BabelNetga integratsiya qilingan.



Turkiy tillar uchun maxsus WordNet tizimi alohida shakllantirilmagan bo'lsa-da, Wikipedia maqolalari, Wikidata obyektlari, turli geografik nomlar va rasmiy onlayn lug'atlar asosida semantik sinsetlar yaratilgan. Shu sababli turkiy tillarda shakllangan sinsetlarning aksariyati ensiklopedik yo'nalishga ega bo'lib, geografik obyektlar, toponimlar, shaxslar, milliy terminlar va madaniy konseptlar tarmoqda ancha keng tasvirlangan. Bu jarayon lingvistik resurslar cheklangan tillar

uchun muayyan bo'shliqni to'ldiradi va ularni global semantik makonning mustaqil ishtirokchisiga aylantiradi. Ayniqsa, o'zbek tilining BabelNet tarkibiga kiritilgani muhim ilmiy hodisa sifatida e'tirof etiladi. Bu o'zbek tilining global semantik tizimda o'z o'rniga ega bo'lishiga, konseptual birliklarning boshqa tillar bilan bog'lanishiga va o'zbek tilidagi ensiklopedik bilimlarning semantik graf tarkibida to'liq aks etishiga imkon yaratadi.

BabelNet sinsetlari faqat til manbalaridan kengayib borish bilan cheklanmaydi; ular so'z turkumlarining semantik vazifalarini ham chuqur aks ettiradi. Tarmoq strukturasi nazar tashlanganda, otlar eng katta semantik qatlamni tashkil etishi tabiiy, chunki konseptual ma'lumotlarning aksariyati aynan otlar orqali ifodalanadi: predmetlar, obyektlar, joylar, shaxslar va abstrakt tushunchalar semantik xaritaning asosiy bo'g'inidir.

Fe'llar nisbatan kamroq bo'lsa-da, ular semantik jarayonlarning dinamik tomonini ifodalaydi. Harakat, jarayon yoki holatga doir fe'l sinsetlar sabab-oqibat bog'lanishlari, vaqt doiralari va jarayonlar zanjirini modellashtirishda alohida ahamiyatga ega.

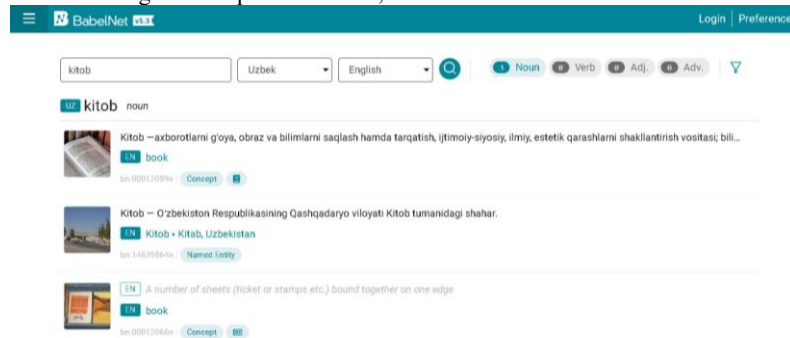
Sifatlar esa konseptlarning turli belgi va xususiyatlarini ko'rsatadi. Ular odatda otlarga bog'lanib keladi. Ravishlar harakatning tarzini ifodalab, fe'llar bilan bog'liq birliklarning semantik kengayishini ta'minlaydi. Biroq ravishlar leksik tizimda nisbatan kam markaziy o'rinni egallaydi, shuning uchun ular tarmoqda kichikroq guruh sifatida ko'rinadi.

BabelNetning konseptual arxitekturasi turli tillarda sinsetlarning qanday shakllangani haqida ham aniq tasavvur beradi. Ingliz, fransuz, ispan, nemis va niderland tillari konseptual boylik bo'yicha yetakchi hisoblanadi. Bunga ushbu tillarda mavjud bo'lgan Wikipedia, Wikidata va boshqa katta bilim omborlarining muntazam yangilanib borishi sabab bo'ladi. Natijada bu tillarda shakllangan konseptlar soni ham,

ularning semantik murakkabligi ham nihoyatda yuqori. Ingliz tili ayniqsa markaziy rol o'ynaydi, chunki global ilmiy adabiyotlar, ensiklopediyalar va rasmiy ma'lumot bazalarining katta qismi aynan ingliz tilida yaratilgan.

BabelNetning eng muhim jihatlardan biri unda leksik-semantik sinsetlar bilan bir qatorda nomlangan obyektlarning (Named Entities) ham bir xil semantik struktura asosida qayta ishlanishidir. Nomlangan obyektlar orasida davlatlar, shaharlar, toponimlar, tarixiy shaxslar, tashkilotlar va madaniy obyektlar mavjud bo'lib, ular til birliklari bilan bir qatorda tarmoqning markaziy qatlami sifatida namoyon bo'ladi. Bu esa BabelNetni faqat leksik model emas, balki to'laonli bilim grafi sifatida ko'rishga imkon beradi.

Tizimdagi sinsetlar ikki turga bo'linadi: tushunchalar (concepts) va nomlangan obyektlar (Named Entities). Masalan, foydalanuvchi "Kitob" so'zini qidiruvga kiritganida, BabelNet uning o'zbek tilidagi shaklini Qashqadaryodagi Kitob shahri sifatida avtomatik tasniflab, toponimga tegishli sinsetni chiqaradi. Bu tasnif GeoNames, Wikidata va Wikipedia kabi manbalar bilan bog'langan holda shakllangan murakkab integratsiya natijasidir. Sinset sahifasida shahar haqidagi geografik joylashuv, koordinatalar, ma'muriy birlik, balandlik va aholi soniga oid ma'lumotlar xarita ko'rinishida interaktiv tarzda berilishi BabelNetning vizual-semantik imkoniyatlarini yanada yaqqol namoyon etadi (2-rasmga qarang).



## 2-rasm. Kitob so'zining qidiruv natijasi

Mazkur misol BabelNetning nomlangan obyektlar bilan ishlashda bir nechta strategik bosqichni amalga oshirganini ko'rsatadi. Avvalo, tizim tilga bog'liq bo'lmagan konseptual identifikator yaratadi, ya'ni "Kitob" toponimi barcha tillar uchun yagona konsept sifatida belgilab olinadi. Keyin u turli tillardagi nomlarini, leksik shakllarini va tarjimalarini o'sha bir sinset ostida jamlaydi. Shundan so'ng, ensiklopedik manbalardan olingan faktik ma'lumotlarni – masalan, aholi soni, geografik joylashuvi yoki ma'muriy bo'linmasini – sinset tarkibiga kiritib, real obyektning ko'p manbadan tasdiqlangan semantik modelini yaratadi. Natijada "Kitob" kabi shahar nomi BabelNetda oddiy tarjima sifatida emas, balki to'liq ma'lumotga ega bo'lgan mustaqil semantik birlik sifatida mavjud bo'ladi. Agar so'z shaxs nomi bo'lsa tizim uniyam ajratib beradi. Masalan, lola so'zi qidiruvga berilganda uning gul nomi ekanligini (tushuncha) va shax nomi ekanligini (nomlangan obyekt) ajratgan. Tizim o'zbek tilida faqat ot turkumiga oid so'zlarning ma'nolarini yaxshi ajrata oladi. Lekin fe'l, sifat, ravish turkumiga oid so'zlarning faqat bir ma'nosini to'g'ri bergan ayrim so'zlarning, asosan murakkab so'zlarning ma'nolarini xato bergan (muzlatgich, sotib olmoq kabi). Masalan, fe'l so'z turkumiga oid yugurmoq so'zi qidiruvga berilganda

BabelNetning tuzilishi nafaqat lingvistik va ensiklopedik bilimlarni integratsiya qiluvchi yirik semantik tarmoq sifatida, balki turli xil amaliy va ilmiy jarayonlarda foydalanishga mo'ljallangan universal platforma sifatida

ahamiyatlidir. Resursning ochiq veb-interfeysi va dasturiy APIlari tadqiqotchilarga semantik tarmoq elementlarini turli ko'rinishlarda o'rganish, tahlil qilish va qo'llash imkoniyatini beradi. BabelNet konseptlarni ko'p tilli sinsetlar shaklida ifodalaganligi bois, foydalanuvchi birgina so'rov orqali ma'lum bir konseptning turli tillardagi shakllarini, glosslarini, misollarini va semantik bog'lanishlarini ko'rishi mumkin. Bu yondashuv uni ko'p tilli semantik tahlil, tarjima texnologiyalari, ma'no aniqlash va ma'lumot qidiruvi kabi yo'nalishlarda asosiy vositaga aylantiradi.

BabelNetning yana bir muhim imkoniyati – domenlar (sohalar) bo'yicha tasniflash. Har bir sinsetga ma'lum bir bilim sohasi (masalan: tibbiyot, biologiya, texnologiya, san'at, tarix) birlashtirilgan bo'lib, bu sinsetning semantik kontekstini aniqlashga yordam beradi. Domen etiketlari ilmiy tadqiqotlarda terminologik qatlamni ajratish, terminlar o'rtasidagi munosabatlarni tahlil qilish, shuningdek, muayyan soha bo'yicha semantik xaritalar tuzishda qo'l keladi.

BabelNet interfeysi semantik graf ko'rinishini ham taklif etadi, bunda konseptlar tugun sifatida, ularning o'zaro munosabatlari esa yo'lar sifatida aks ettiriladi. Bu vizual klasterlarni, tematik guruhlarini va konseptlarning joylashuv darajasini intuitiv tarzda ko'rish imkonini beradi. Graf orqali foydalanuvchi sinset atrofidagi so'zning semantik maydonini chuqurroq anglaydi.

