

WordNet – ONLINE LEKSIK MA'LUMOTLAR BAZASI

Ravshanov Shuhratjon Ibragim o'g'li

Mirzo Ulug'bek nomidagi

O'zbekiston Milliy universiteti magistranti

ravshanovshuhrat0103@gmail.com

Annotatsiya. WordNet - ingliz tilining katta leksik ma'lumotlar bazasi. Otlar, fe'llar, sifatlar va qo'shimchalarni kognitiv sinonimlar to'plamiga (sinsetlarga) birlashtirilib, ularning har biri alohida tushunchani ifodalaydi. Sinsetlar konseptual-semantik va leksik munosabatlar yordamida o'zaro bog'liqdir. Natijada paydo bo'lgan mazmunli so'zlar va tushunchalar tarmog'i brauzer yordamida boshqariladi. WordNet shuningdek, yuklab olish uchun ommaga ochiq taqdim etiladi. WordNet tuzilmasi komputer lingvistikasi va tabiiy tillar jarayoni uchun foydali vosita hisoblanadi.

Kalit so'zlar: *WordNet, sinset, tezaurus, giperonimiya, giponimiya, meronimiya, POS aloqalar, psixolingvistika*

WordNet - ONLINE LEXICAL DATABASE

Annotation. WordNet is a large lexical database of English. Nouns, verbs, adjectives, and suffixes are grouped into a set of cognitive synonyms (syns), each of which represents a separate concept. Synsets are interconnected using conceptual-semantic and lexical relationships. The resulting network of meaningful words and concepts is managed using a browser. WordNet is also open to the public for download. The WordNet structure is a useful tool for the process of computer linguistics and natural languages.

Key words: *WordNet, sinset, thesaurus, hyperonymy, hyponymy, meronymy, POS communication psycholinguistics*

Kirish. WordNet - 200 dan ortiq tillarda qo'llaniladigan so'zlar o'rtasidagi semantik munosabatlarning leksik ma'lumotlar bazasi. WordNet so'zlarni semantik munosabatlarga, shu jumladan sinonimlar, giponimlar va meronimlarga bog'laydi. Sinonimlar qisqa ta'riflar va ularga keltirilgan misollardan foydalanib, sinsetlar guruhlariga ajratiladi. Shu sababdan, WordNet lug'at va tezaurus birikmasi va kengaytmasi sifatida qaralishi mumkin. U veb-brauzer orqali foydalanuvchilarga taqdim etilgan bo'lsa-da, u, asosan, avtomatik matn tahlili va sun'iy intellekt dasturlarida ko'p qo'llaniladi.

WordNet dastlab ingliz tilida Prinston universiteti Ilm-fan laboratoriyasida 1985 yilda psixologiya professori Jorj Armitaj Miller rahbarligida yaratilgan. WordNet dasturi o'ziga xos psixolingvistik jihatlarga asoslangandir. WordNet loyihasining dastlabki maqsadi 60-yillarning oxirida ishlab chiqilgan insonning semantik xotirasi nazariyalariga mos keladigan leksik ma'lumotlar bazasini yaratish edi. Psixologik tajribalar shuni ko'rsatdiki, ma'ruzachilar o'z fikrlarini iyerarxik shaklda tashkil etganlar. Kontseptual bilimlarni olish uchun talab qilinadigan qidirish vaqti spikerning bilimga kirish uchun "o'tishi" kerak bo'lgan ierarxiya soni bilan bevosita bog'liq bo'lgan. Shunday qilib, ma'ruzachilar kanareykalar qo'shiq aytishini tezroq tekshirib ko'rishlari mumkin edi, chunki kanareyka - bu kuylovchi qush, ammo kanareykalar ucha olishini tekshirish uchun biroz ko'proq vaqt kerak edi (bu yerda ular "qush" tushunchasidan yuqoriroq darajaga bog'lanishlari kerak edi) va undan ham ko'proq vaqt kanareykalarning terisi borligini tekshirib ko'rishga sarflandi (chunki bu "hayvon" ga qadar ko'p darajadagi giponimiya bo'yicha qidiruvni talab qilardi). Bunday psixolingvistik tajribalar va uning negizidagi nazariyalar tanqidga uchragan bo'lsa-da, WordNet tashkiloti o'zining eksperimental dalillariga suvandi. Masalan, anomik afaziya kasalligida bemorlarning faqat ma'lum bir semantik toifadagi so'zlarni ishlab chiqarish qobiliyatiga ta'sir qilishi aniqlandi. Va bu ularga inson har bir tushunchani o'z xotirasida ma'lum bir fayllarda ketma-ketlikda saqlashiga dalil bo'ldi.

Tuzilishi. WordNet-dagi asosiy munosabat, berkitish va yopish yoki mashina va avtomobil so'zlari orasidagi sinonimlikdir. Sinonimlar - bir xil tushunchani bildiruvchi va ko'p kontekstda bir-birining o'rnini bosadigan so'zlar bo'lib, ular tartiblanmagan to'plamlarga (sinsets) birlashtirilgan. WordNetning 117 000 ta sinsetining har biri "kontseptual munosabatlar" yordamida boshqa sinsetlar bilan bog'langan. Bundan tashqari, sinset qisqacha ta'rifga (gloss) ega bo'lib, aksariyat hollarda bir yoki bir nechta qisqa jumlar bilan izohlanadi. Bir necha ma'noli so'z shakllari bir necha xil sinsetlarda kuzatilishi mumkin. Shu sababli, WordNetdagi har bir shakl-ma'no juftligi o'ziga xosdir.

Munosabatlar. Sinsetlar orasida eng ko'p qo'llangan munosabatlar ikkinchi darajali munosabatlar (shuningdek, giperonimiya, giponimiya yoki iyerarxik munosabatlar) deb ataladi. Ular umumiy sinsetlarni (mebel, uy jihozi) tobora aniqroq shakllangan sinsetlarga (krovat, yotoq) bog'laydi. Shu sababli, WordNet ushbu mebellar kategoriyasiga krovat va yotoq kiradi; yoki aksincha, krovat va yotoq kabi tushunchalar mebellar kategoriyasini tashkil qiladi, deb qabul qiladi.

Barcha ot iyerarxiyalari oxir-oqibat ildiz tuguniga borib taqaladi. Giponimiya munosabati o'timli bo'lib, agar kreslo stulning bir turi bo'lsa, va stul - bu mebelning bir turi bo'lsa, unda kreslo - bu mebelning bir turi deb qaraladi. WordNet turdosh otlarni va atoqli otlarni (ma'lum bir shaxslar, mamlakatlar va geografik hududlar) ajratib turadi. Barg o'simlik tanasining bir qismiga, Albert Eynshteyn esa olim so'ziga misol bo'la oladi. Atoqli otlar o'z iyerarxiyalarida eng so'nggi o'rinda yoziladi.

Meronimiya stul va suyanchiq, o'rindiqlar va oyoqlar kabi sinsetlar o'rtasida bo'ladigan butun va qismning aloqasidir. Qismlar o'zlariga butundan har bir narsani qabul qilib oladi: agar stulning oyoqlari bo'lsa, unda suyanchiqning ham oyoqlari bor. Ammo butun hech qachon qismdan biror xususiyatni qabul qilib olmaydi, chunki ular umumiy sinfga emas, balki faqat o'z turiga xos bo'lishi mumkin. Masalan, stullar va suyanchiqlarning oyoqlari bor, ammo bu barcha mebellarning oyoqlari bor degani emas.

Fe'l sintezlari ham iyerarxiyalarga joylashtirilgandir. Daraxtning eng tubida joylashgan fe'llar (troponimlar) biror hodisani o'ziga xos uslubda tavsiflab beradi, masalan, suhbat qurmoq – gaplashmoq - pichirlashmoq. Belgilangan uslub semantik maydonga bog'liq bo'ladi; Hajm fe'llarni batafsil ishlab chiqishning bir o'lchovi xolos. Boshqa o'lchovlar tezlik (siljish-yurish-yugurish) yoki tuyg'ular kuchliligi (yoqtirmoq-sevmoq-maftun bo'lmoq) bo'lishi mumkin. Bir-biriga doimiy va bir tomonlama ta'sir ko'rsatadigan, hodisalarni tavsiflovchi fe'llar o'zaro bog'langandir. Masalan, sotmoq – sotib olmoq, ko'rmoq - ko'rsatmoq, erishmoq – harakat qilmoq.

Sifatlar esa antonimlik jihatidan tartiblangandir. Nam va quruq, yosh va qari kabi "to'g'ridan-to'g'ri" antonim juftliklari bir-birlari o'rtasida kuchli semantik tafovutni aks ettiradi. Ushbu qarama-qarshi sifatning har biri o'z navbatida bir qator "semantik jihatdan o'xshash" so'zlar bilan bog'lanadi: quruq so'zi qurg'oq, qurg'oqchil, qurigan, yog'insiz so'zlariga va nam so'zi ho'l, shilta, shalabbo, seryog'in so'zlariga bog'langan. WordNet dasturida har bir antonim juftliklarning o'ziga semantik jihatdan o'xshash sifatlarini ham bir-biriga antonim deb qabul qilib boramiz.

Gap bo'laklari orasidagi aloqalar. WordNet aloqalarining aksariyati bir xil gap bo'laklarini (POS) bog'laydi. WordNet, aslida, to'rtta kichik to'rdan iborat bo'lib, ularning har biri otlar, fe'llar, sifat va ravishlar orasidagi o'zaro gap bo'laklarni aniqlovchidan iborat. Bu munosabatlarga "morfosemantik" aloqalardan tashkil topib, ular bir xil ma'noga ega bo'lgan, o'zakni birlashtiradigan, semantik

jihatdan o'xshash so'zlar orasida qo'llaniladi: kuzatmoq (fe'l), kuzatuvchi, kuzatuv (otlar). Ko'plab ot-fe'l juftliklarida fe'lga nisbatan otning semantik roli ajratib ko'rsatiladi: kuzatuvchi – kuzatmoq, sevgi – sevmok, bo'yoq – bo'yamoq.

Xulosa. WordNet yuzaki ravishda tezaurusga o'xshaydi, chunki u so'zlarni ma'nolariga qarab birlashtiradi. Biroq, ba'zi bir boshqa manbalarda qo'llanilmagan farqlar mavjud. Birinchidan, WordNet nafaqat so'z shakllari - harflar qatorini, balki so'zlarning o'ziga xos ma'nolarini o'zaro bog'laydi. Natijada, tarmoqdagi bir-biriga yaqin bo'lgan so'zlar semantik jihatdan ajralib chiqadi. Ikkinchidan, WordNet so'zlar orasidagi semantik munosabatlarni belgilaydi, tezaurusda esa aksincha so'zlarning yig'indisi ma'no o'xshashligidan tashqari boshqa bir aniq qolipga amal qilmaydi.

Foydalanilgan adabiyotlar

1. **Chafe, W.** 1970. *Meaning and the Structure of Language*. Chicago: University of Chicago Press.
2. George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller. 1993. Introduction to WordNet: An On-line Lexical Database.
3. <https://wordnet.princeton.edu/>
4. Razieh Ehsani, Ercan Solak, and Olcay Taner Yildiz, Isik University. Constructing WordNet for Turkish Using Manual and Automatic Annotation.
5. Daniel Jurafsky, James H. Martin. *Speech Language Processing*.