

SEMANTIK TIZIMLAR VA LINGVISTIK ONTOLOGIYALAR

O'ZBEK TILIDA FE'L SO'Z TURKUMI UCHUN VERBNET YARATISH

Sharipov Maqsud Siddiqovich

Texnika fanlari nomzodi, dotsent,

m.sharipov@urdu.uz

UrDU, Kompyuter ilmlari kafedrası dotsenti

Mammatova Mahbuba Davronbek qizi

mammatovamahbuba1996@mail.com

UrDU, Kompyuter ilmlari kafedrası magistranti

Annotatsiya. Tabiiy tilni qayta ishlash masalalari orasida fe'l so'z turkimiga oid so'zlarni qayta ishlash masalalarini yechish muhim o'rin egallaydi. O'zbek tilida fe'llarni stemming va lemmatizatsiya qilish hamda kompyuter tarjima dasturlarini yaratish kabi masalalarni yechish uchun fe'llar uchun VerbNet yaratish masalasi dolzab hisoblanadi. Ushbu ishda o'zbek tili uchun VerbNetning tuzilmasini loyihalash va uni yaratish masalasi ko'rib chiqilgan.

Annotation. Among the problems of natural language processing, solving the problems of processing words related to the word group of verbs occupies an important place. The issue of creating a VerbNet for verbs to solve problems such as stemming and lemmatization of verbs in the Uzbek language and creating computer translation programs is considered important. In this work, the issue of designing and creating the structure of VerbNet for the Uzbek language is considered.

Аннотация. Среди задач обработки естественного языка важное место занимает решение задач обработки слов, относящихся к словесной группе глаголов. Важным считается вопрос создания VerbNet глаголов для решения таких задач, как stemming и лемматизация глаголов в узбекском языке и создание программ компьютерного перевода. В данной работе рассматривается вопрос проектирования и создания структуры VerbNet для узбекского языка.

Kalit so'zlar. *Sinonim, giperonim, giponim, antonim, paronim, verbnet, wordnet.*

Tabiiy tillarni qayta ishlash masalalarini yechish uchun tildagi so'zlarni bir-biriga bog'lovchi ma'lumotlar bazasini yaratish muhim o'rin egallaydi. Shu sababli ushbu masalani yechish uchun WordNet yaratilgan. Wordnet bu so'z mazmunlari asosida yaratilgan juda keng qamrovli lug'atdir. Wordnetning asosiy qurilish bloklari sinsetlar bo'lib, ular sinonim sinset a'zolaridan iborat. O'zbek WordNetiga asos bo'lgan Turk WordNeti strukturasi, foydalanish holati Til manbalaridan so'zlar munosabatlari chiqarib olingan. Dastlabki 1310 ta sinsetlar to'plamining

tarjimasidan so‘ng, turk jamoasi mashina tomonidan o‘qiladigan turk tilining izohli lug‘atidan sinonimlar, antonimlar va giponimlarni avtomatik ravishda chiqarib olishga harakat qilishdi [Madatov, 2022].

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 [Ravshanov, 2021].

WordNet ingliz tilining leksik ma‘lumotlar bazasi hisoblanadi. Undagi ot, fe‘l, sifat va ravish so‘z turkumlari muayyan konseptni ifodalovchi sinonimlar guruhiga (Sinset) birlashgan. Sinsetlar konseptual-semantik va leksik munosabatlarga ko‘ra o‘zaro bog‘langan. WordNetdan foydalanish va uni o‘rnatib olish qulay. Uning tuzilishi kompyuter lingvistikasi va tabiiy tilni qayta ishlash uchun foydali. WordNetda fe‘l so‘z turkumiga kiruvchi troponimlar kiritilgan. Unga ko‘ra u yoki bu fe‘l boshqa muayyan shakliga ko‘ra bog‘lanadi. Masalan: vaysamoq, gapirmoq fe‘lning troponimi hisoblanadi. Shuningdek fe‘llar ham antonimik munosabatda bo‘lishi mumkin turmush qurmoq-ajrashmoq [Abduraxmonova, Haydarov].

VerbNet_- Bu fe‘llarni semantik, sintaktik bog‘lanish harakati asosida guruhlaydigan leksik manba. Biz 5 tomlik o‘zbek tilining izohli lug‘ati tarkibida mavjud 8519 ta fe‘llar bazasini to‘pladik. To‘plangan fe‘l shakllarini tuzilishi va ma‘no guruhi jihatdan belgilab oldik. Ro‘yxatdagi yasama fe‘l shakllarining asosi qaysi turkumga oidligi, qanday qo‘shimcha orqali yasalgani, fellar tarkibidagi mavjud lug‘aviy shakl yasovchi qo‘shimchalarni (nisbat yasovchi qo‘shimchalar) guruhlab tahlil qildik. Bazadagi fe‘llarning mavjud sinonim, antonim, paronim, giponim va giperonimlari to‘plandi. Biz hozirgacha 447 ta sinonimli fe‘l, 91 ta paronimli fe‘l, 86 ta antonimli fe‘l shakllarini to‘pladik. Bundan tashqari to‘plangan fe‘llarni qanchasi tub (faqat asosdan yoki shakl yasovchi qo‘shimchali), qanchasi esa so‘z yasovchi qo‘shimchalar orqali boshqa turkumlardan yasalganligini aniqladik. Va nechta fe‘l shakli qaysi so‘z turkumidan yasalganligining statistikasi chiqarildi. Qaysi qo‘shimcha qaysi turkumdagi so‘zdan fe‘l shakli yasashi va nechta o‘rinda yasaganini hisobladik. Bu ma‘lumotlar 1-jadvalda ko‘rsatib o‘tiladi:

1-jadval

Jami fe‘llar	8518
Tub fe‘llar	3461
Otdan yasalgan fe‘llar	2447
Sifatdan yasalgan fe‘llar	1389

Taqlid soʻzdan yasalgan feʼllar	1143
Ravishdan yasalgan feʼllar	35
Sondan yasalgan feʼllar	17
Olmoshdan yasalgan feʼllar	17
Modal soʻzdan yasalgan feʼllar	8

Yuqoridagi maʼlumotlarga tayangan holda har bir turkumdan qaysi qoʻshimcha orqali nechta feʼl shakli hosil qilinganiga asoslanib eng koʻp feʼl shakli yasagan qoʻshimchalarni jadval orqali berib oʻtamiz.

2-jadval

la	3390ta
illa	414ta
lash	410ta
lan	331ta
a	24ta
ulla	60ta

SINONIM - bir umumiy maʼnoga ega boʻlgan qoʻshimcha maʼnosi uslubiy va boshqa munosabat kabilarni ifodalovchi xususiyatlari bilan oʻzaro farqlanadigan til birliklari. Misol uchun: kulmoq, jilmaymoq, tirjaymoq, ishshaymoq. Oʻzaro sinonim boʻlgan soʻzlar guruhi sinonimik qator deyiladi. Yuqoridagi sinonimik qatordagi soʻzlar bir umumiy “kulish” manosi ega, ammo “kulmoq” soʻzi va “tirjaymoq”, “ishshaymoq” soʻzlari bir-biridan uslubiy jihatdan farqlanadi.

GIPERONIM - jins belgisini bildirgan predmetning nomini ifodalovchi koʻpgina maʼnoni semantik jihatdan umumlashtiruvchi mikrotizimning markaziy leksemasi, dominantasi sifatida namoyon boʻluvchi lugʻaviy birlik. Misol uchun: kulmoq, jilmaymoq, tirjaymoq, ishshaymoq qatorida dominanta soʻz “kulmoq” giperonim.

GIPONIM - maʼlum jins turining nomini hamda oʻzining semantik tarkibida implist tarzida jins maʼnosini ham ifodalovchi, semantik jihatdan giperonimga nisbatan boy boʻlgan lugʻaviy birlik. Misol uchun: kulmoq-giperonim; jilmaymoq, tirjaymoq, ishshaymoq-giponim. Feʼl maʼnoviy guruhlarida ham markaz va qurshov leksemalari farqlanadi. Masalan, maʼnodosh feʼllarda bosh (dominanta) soʻz, uyadoshlik paradigmasida uya (giperonim) markaz leksemasi hisoblansa, boshqa maʼnodosh va uyadosh (giponimlar) qurshov leksemasi deyiladi. [Kulmoq] feʼli dominanta sifatida bunga maʼnodosh boshqa leksemalarni, [buzmoq] esa buzishning turli koʻrinishini ifodalaydigan giponim leksemani uyushtirib turadi. Soʻzlar yirikroq butunliklarga birlashtirilganda, markaz leksemasigina bu butunlikdan joy oladi. Qurshov leksemalari esa uning “soyasi” sifatida quyida qolaveradi. Masalan, [kulmoq], [yigʻlamoq], [xursand boʻlmoq] leksemalari oʻz atrofida koʻplab feʼllarni birlashtirgan. Lekin ular bu uch leksema mansub tizimga kira olmavdi, quyi bosqichda qoladi [Sayfullayeva va b., 2009].

ANTONIM-bir-biriga qarama-qarshi ma'nolarni ifodalovchi so'zlar juftligi. Antonimik juft hosil bo'lishi uchun ikkita mustaqil tushuncha ma'no jihatdan o'zaro qarama-qarshi bo'lishi kerak. Fellardagi bo'lishli-bo'lishsizlik xususiyati antonimlikni vujudga keltirmaydi. Misol uchun: keldi-ketdi, maqtamoq-qoralamoq, kirmoq-chiqmoq.

PARONIM - talaffuzi, eshitalishi va morfem tarkibi o'xshash, leksik ma'nolari boshqa-boshqa yoki qisman yaqin bo'lgan so'zlar. Misol uchun: alishmoq-olishmoq, artmoq-archmoq, chuchimoq-cho'chimoq.

Foydalanilgan adabiyotlar:

1. Abdurakhmonova N., Khaydarov M. (2019). Ўзбек тилида wordnet яратиш масалаларига доир. Foreign Languages in Uzbekistan. 19-27. 10.36078/1570167968.
2. Abjalova M, Rashidova U, Abdumuminov B. Matnlarni qayta ishlash manbalari. Computer linguistics: problems, solutions, prospects 2022;
3. Akhmedovich Khabibulla. (2022). Creating of the Uzbek WordNet based on Turkish WordNet. AIP Conference Proceedings. 2432. 10.1063/5.0089532.
4. Boltayevich E.B, Adali E, Mirdjonovna Sh.X, Xolmo'minovna A.O, Yuldashevna X.Z, Nizomaddin Uktamboi o'g'li X. The Problem of Pos Tagging and Stemming for Agglutinative Languages (Turkish, Uyghur, Uzbek Languages). UBMK 2023 - Proceedings: 8th International Conference on Computer Science and Engineering, 2023, 57 – 62.
5. Ravshanov Shuhratjon. (2021). Wordnet-online leksik ma'lumotlar bazasi. Computer linguistics: problems, solutions, prospect. 1. 149-152. 10.52773/cl:psp/vol_1_issue_1/a30.
6. Sayfullayeva R.R., Mengliyev B.R., Boqiyeva G.H., Qurbonova M.M., Yunusova Z.Q., Abuzalova M.Q. Hozirgi o'zbek adabiy tili. O'quv qo'llanma. - T., 2009. - 416 b.
7. Sharipov M, Kuriyozov E, Yuldashev O, Sobirov O. UzbekTagger: The rule-based POS tagger for Uzbek language. arXiv preprint arXiv:230112711 2023;
8. Sharipov M, Mavlonova M, Sobirov O. O'zbek tilidagi matnlarda ot va olmosh so'z turkumlariga tegishli so'zlarni aniqlovchi algoritm yaratish. Computer Linguistics: Problems, Solutions, Prospects 2023; 1.
9. Sharipov M, Salaev U, Matlatipov G. Implemented stemming algorithms based on finite state machine for uzbek verbs| computer linguistics: problems. Solutions, prospects 2021;
10. Sharipov M, Sobirov O. Development of a Rule-Based Lemmatization Algorithm Through Finite State Machine for Uzbek Language. CEUR Workshop Proceedings, CEUR-WS 2022, 154 – 159.
11. Sharipov M, Yuldashov O. UzbekStemmer: Development of a Rule-Based Stemming Algorithm for Uzbek Language. CEUR Workshop Proceedings, CEUR-WS 2022, 137 – 144.