

KORPUS LINGVISTIKASI

Uznet lingvistik ontologiyasini yaratishni loyihalash konsepsiyasi

Manzura Abjalova¹

Abstrakt

Jahonda inson faoliyatining bilim talab qiladigan zamonaviy sohalari rivojlanishi kompyuter texnologiyalari rolining oshishi bilan belgilanadi. Bugungi kunda axborot oqimi sezilarli darajada ko'paymoqda, endi ularni saqlash, taqdim etish, formallashtirish va tartibga solish, shuningdek, avtomatik qayta ishlashning yangi usullarini izlash zarurati yuzaga kelmoqda. Shu bois turli amaliy maqsadlarda qo'llanilishi mumkin bo'lgan keng qamrovli bilim bazalariga qiziqish ortib bormoqda. Ayniqsa, inson omilisiz matndan har qanday ma'lumotni chiqarib oluvchi neyro-to'rlarga asoslangan tizimlarga ehtiyoj katta. XX asr yarimida butunjahon tarmog'i bilan bir qatorda semantik veb'lar paydo bo'ldi, unda gipermatnli sahifalardagi elementlar semantikasi haqida axborot tashuvchi qo'shimcha teglar bilan ta'minlandi. Semantik vebning ajralmas qismi ontologiya tushunchasi bo'lib, u so'zlar tarmog'idan iborat leksik ma'lumotlar bazasi hisoblanadi.

Mazkur maqolada o'zbek tili ontologiyasi – UzNet tizimini yaratish konsepsiyasi, uni loyihalash asoslari, UzNet ontologiyasining biznes-jarayon modeli ko'rsatildi.

Kalit so'zlar: *lingvistik ontologiya, semantik veb, UzNet, semantik munosabatlar, konsepsiya, loyihalash.*

Kirish

XX asrda dunyo amaliy tilshunosligi va kompyuter lingvistikasida axborot qidiruv imkoniyatini oshirish, avtomatik tarjima tizimlarini takomillashtirish, matnlarni komponent tahlil qilishga erishish, tilning lug'at boyligini aks ettirish maqsadida

¹Abjalova Manzura Abdurashetovna – filologiya fanlari bo'yicha falsafa doktori (PhD), Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti Kompyuter lingvistikasi va raqamli texnologiyalar kafedrasi dotsent v.b.

E-pochta: abjalovamanzura@navoiy-uni.uz

ORCID: 0000-0002-1927-2669

Iqtibos uchun: Abjalova, M.A. 2022. "UzNET lingvistik ontologiyasini yaratishni loyihalash konsepsiyasi". *O'zbekiston: til va madaniyat. Amaliy filologiya* 1(5): 123-139.

yaratilgan lingvistik ontologiyalar tezauruslardan tildagi semantik munosabatlarning o'zaro tarmoqlanishi bilan ajralib turdi. Natijada WordNet ingliz tili leksik ma'lumotlar bazasi asosida boshqa ko'plab tillar resurslarini yaratish keng tus oldi. Shu sababli tabiiy tilni qayta ishlash (NLP), tilni lingvistik modellash, so'z turkumlarini teglash, tilda semantik munosabatlar to'plami (sinonimiya, meronimiya, giperonimiya, antonimiya)ni shakllantirishga alohida e'tibor qaratildi.

Bugungi kunda jahon miqyosida ko'plab semantik tizimlar yaratilgan bo'lsa-da, ammo lingvistik ontologiya va tezauruslarning sanoqlisi ularga qo'yilgan talablarga javob beradi. Lingvistik ontologiyalarni yaratish an'anasi WordNet asosidagi lingvistik resurslar ta'minotini o'rganish va yaratishga yo'naltirilgan harakatlar natijasida yuzaga kelgan.

Lingvistik ontologiyani yaratish va qo'llash bo'yicha ko'plab ishlar Rossiyada hamda xorijda amaliy tilshunoslik sohasida olib borilgan va hozirda jadal ravishda davom etmoqda. Ushbu tadqiqotda J.Miller va K.Fellbaum P. Butelaar, F.Chimiano, P.Haaza, B.V.Dobrov, N.V.Lukashevich, O.Narinyan, I.Azarova tadqiqotlari ahamiyatli sanaladi.

"O'zbek tili ontologiyasini yaratish konsepsiyasi" yirik lingvistik ontologiyalar – SUMO, OMEGA, DOLCE, Princeton WordNet, KeNet, RuThez va RussNet kabi ontologiya hamda tezauruslarni o'rganish, tahlil qilish natijada o'zbek tili ontologiyasini yaratish bo'yicha metodologik bilimlar, muayyan sohaga xos bo'lgan ma'lumotlarni qayta ishlashga imkon beruvchi tajribalar asosida ishlab chiqildi.

Asosiy qism

Lingvistik ontologiya yoxud *til ontologiyasi* terminlari tilshunoslik faniga qaraganda, axborot texnologiyalari sohasida ko'proq qo'llaniladi, asosan, matnlarni avtomatik tarzda qayta ishlash uchun ixtisoslashtirilgan axborot qidirish tezaurusi, ya'ni tilning lug'at boyligini o'zida jamlagan, so'zlarning semantik munosabatlari o'rnatilgan (yoxud so'zlar tarmog'iga ega) turli maxsus lingvistik dasturiy ta'minotlarni anglatadi. Lingvistik ontologiya (LO) til borlig'i va mohiyati haqidagi fan sifatida kamdan-kam tilga olinadi. LO lisoniy borliqni tahlil qilish orqali tilning mohiyatini ochib berishga xizmat qiladi.

Ontologiyalar ko'plab kompyuter dasturlari uchun [Eiji Aramaki] ma'lumot manbasi sifatida qo'llaniladi (axborot qidirish,

matnni tahlil qilish, avtomatik tarjima, bilimlarni yig'ish va boshqa axborot texnologiyalari uchun). Ontologiya murakkab va xilma-xil ma'lumotlarni samarali qayta ishlashda yordam beradi [Гладун]. Ma'lumotlar bilan ishlashning bunday usuli dasturlar uchun insonga tushunarli bo'lgan, ammo kompyuterga ma'lum bo'lmagan semantik farqlarni tanib olishga imkon beradi.

Zamonaviy axborot-qidiruv ilovalari keng sohalarda ishlaydi.

So'nggi 15 yildan beri ontologiyalar sun'iy intellekt sohasining diqqat markazida bo'lib kelmoqda. Ular yaratiladi, qo'llaniladi va o'rganiladi. Ontologiyalardan foydalanish holatlari [Клещев]:

- axborotlarni tashuvchi terminlarning ma'nosini aniqlashtirish;

- axborotning qayta qo'llanilishini ta'minlash;

- ma'lumotlarni tahrirlashni nazorat qilish (dasturiy vositada boshqaruvchi ma'lumotlar sifatida);

- tahlilchi faoliyatining bir qismi sifatida dasturiy vositalarni (shu jumladan, bilimga asoslangan tizimlarni) ishlab chiqishda foydalanish.

T.Gavrilovanning fikriga ko'ra, ontologiyalar bilimlarni boshqarish tizimlari, shuningdek, ko'p agentli intellektual tizimlar sohasida faol qo'llaniladi [Гаврилова]. "Ontologiya asosida samarali qidiruvni tashkil etish" nomli maqolada ta'kidlanganidek, ontologiyalar dasturiy ta'minot tizimlari uchun bilimlar bazasi komponentlarining qurilish bloklari sifatida foydalanishi mumkin, shuningdek, obyektga yo'naltirilgan tizimlardagi obyekt diagrammalari, ma'lumotlar bazasining kontseptual sxemalarini qurish bloklari sifatida foydalanish mumkin. Ontologiyalar yordamida turli xil intellektual tizimlar, xususan, ekspert tizimlaridagi muammolarni yechish uchun bilim bazalarini shakllantirish mumkin [Почеева]. O'zbek tilini Internet tiliga aylantirish, uning xalqaro nufuzini oshirish, davlat tili sifatida maqomini yanada mustahkamlash maqsadida O'zbek tili Milliy korpusi bilan birgalikda O'zbek tili ontologiyasini ham yaratish o'ta muhim vazifa sanaladi.

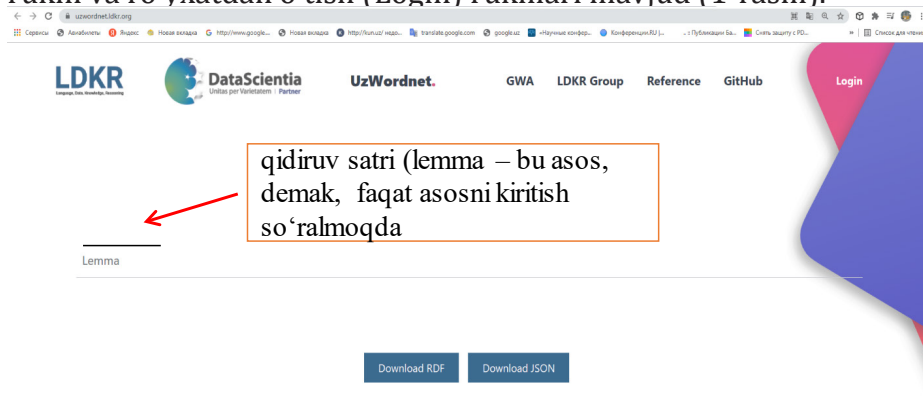
Agar biz o'zbek tilida ma'lumot qidirish dasturlarida foydalanish uchun lingvistik va terminologik resurslarni yaratmoqchi bo'lsak, u holda bu resurslar o'zbek tili leksik sathini keng qamrab olishi, shuningdek, hujjatlar va so'rovlarni qayta ishlashning avtomatik rejimlaridan foydalanish imkoniyatiga ega bo'lishi kerak.

Tadqiqotning oldingi bo'limlarida turli lingvistik resurs va ontologik tizimlar ko'rib chiqildi. Jumladan, lingvistik ontologiyalar, tezauslar, axborot-qidiruv tezauslari, vordnet tipidagi lingvistik

resurlarni yaratish texnologiyalari tadqiq etildi. Ta'kidlash joizki, ushbu manbalarni, ayniqsa, til ontologiyasini yaratish o'ta mashaqqatli va og'ir jarayon hisoblanadi. Albatta, ularning barchasidan ma'lumot qidirish muammolarini hal qilish jarayonida manba sifatida foydalanishda ba'zi muammolar uchraydi. Jumladan, tipologik jihatdan tillarning turli oilalarga mansubligi, tillar flektiv va agglutinatib tabiatga ega ekanligi ingliz tili leksik ma'lumotlari bazasi – WordNet tizimidagi ma'noviy munosabatlar aynan muvofiq kelmaydi. So'z turkumlari tegini belgilashda ishliz tilidagi ravish turkumi o'zbek tilida sifat turkumidagi so'z tarjimasiga muvofiq kelishi mumkin.

Xalqaro Wordnet assotsiatsiyasi saytida berilgan ma'lumotlar [<http://globalwordnet.org/resources/wordnets-in-the-world/>] asosida tayyorlangan tahliliy ma'lumotlar berildi. <http://globalwordnet.org/> Saytning "Wordnets in the World" (Wordnetlar dunyoda) sahifasida Princeton Wordnet tizimi zamirida o'nlab dunyo tillari uchun yaratilgan 78 ta Wordnet keltirilgan.

Ro'yxatning oxirida "(Shimoliy) o'zbek tiliga asoslangan" UzWordnet resursi o'rin olgan va u ko'p tilli ochiq Wordnet formati (the Open Multilingual Wordnet (Bond and Paik, 2012; Bond and Foster, 2013)) [<http://compling.hss.ntu.edu.sg/omw/>] asosida ishlab chiqilgan. Mazkur resurs sahifasi [<https://uzwordnet.ldkr.org/>]da 6 ta, ya'ni UzWordnet so'z qidiruv sahifasi (UzWordnet), xalqaro Wordnet assotsiatsiyasi (XWA) saytiga havola, guruh tarkibi haqida ma'lumot (LDKR Group), UzWordnet to'g'risidagi maqolaga havola (Reference), GitHub, ya'ni ushbu resurs formati, yaratuvchilari, loyiha rahbari, 1 yillik litsenziyasi haqida ma'lumotlar rukni va ro'yxatdan o'tish (Login) ruknlari mavjud (1-rasm).



1-rasm. *UzWordnet* resursining qidiruv oynasi

Loyiha Toshkent shahridagi Inha universitetida amalga oshirilgan, loyiha rahbari – Alessandro Agostini. Manbaning “GitHub” ruknida shunday yozilgan: “UzWordnet – bu (shimoliy) o‘zbek tili (ona tilida: *o‘zbek tili*) uchun Princeton Wordnet bilan mos keladigan (prototip versiya) [Agostini A., Usmanov T. and other] leksik-semantik ma’lumotlar bazasi yoki “so‘zlar tarmog‘i”. Biz o‘zbek tilining hisoblash jihatlarini o‘rganish va umuman, (shimoliy) o‘zbek tilidan IT ilovalari va Internetda foydalanish imkoniyatlari uchun ma’lumotlar bazasi va tamoyillarini qo‘llash, bilimlar grafigi tamoyillari hamda usullarini rag‘batlantirish, qo‘llab-quvvatlash hamda kengaytirishga harakat qildik va uni ochiq manbali kod sifatida taqdim etdik.

(Shimoliy) o‘zbek tili O‘zbekistonda rasmiy davlat tili hisoblanadi. Bu til turkiy til bo‘lib, unda dunyo bo‘ylab 26,8 million kishi, shu jumladan, chet elda yashovchi etnik o‘zbeklar ham gaplashadi” [<https://github.com/LDKR-Group/UzWordnet>].

Tanqidiy mulohazalarni yuzaga keltirishga sabab bo‘lgan ushbu izohlar va UzWordnet resursining amaldagi qidiruv natijalari asosida quyidagi fikrlarni bildirish mumkin:

1. UzWordnet resursini yaratishda shimoliy o‘zbek tiliga asoslanilgani aytiladi. Vaholanki, o‘zbek adabiy tilining mintaqaviy ko‘rinishlari mavjud emas va lingvistik ontologiyada muayyan tilning butun lisoniy ko‘rinishi aks etishi birlamchi talab hisoblanadi. Shu bois lingvistik ontologiyalar tarmoqlangan leksik ma’lumotlar tizimi hisoblanadi.

2. Bir qancha so‘zlar qidiruvi natijasida ayon bo‘ldiki, UzWordnet resursi Princeton Wordnet tizimidagi birliklar va ularning izohini berish bilan cheklanadi. Buni guruh tarkibi ham dalillaydi: *“UZwordnet tuzishning asosiy jarayoni – bu PWN (3.0 versiya)ning o‘zbek tiliga avtomatik tarjimasini hisoblanadi ... (Tarjima uchun) leksik manba sifatida biz Sh.Butayev va A.Irisqulovning “Inglizcha-o‘zbekcha lug‘at”idan (2008) foydalanganmiz [Agostini A., Usmanov T. and other].*

3. UzWordNet jamoasi mazkur resursining ishlash algoritmi PWN’ga asoslanganini qayd etadi va PWN bilan ulanishni tiklash algoritmi sifatida quyidagi algoritmni taqdim qiladi:

Input: S , a data.pos file from Princeton WordNet (PWN, v3.0)

Input : D , English-Uzbek dictionary in tabular form for a specific PoS

Output: W , the UzWordNet (UzW, v1.0)

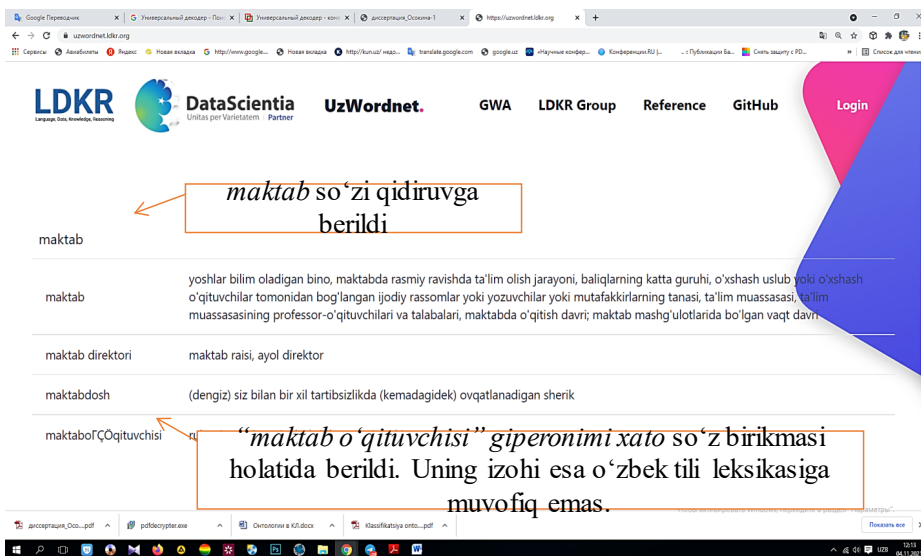
1 $W \leftarrow \emptyset$

```

2 for each synset  $\in S$  do
3   for each lemma  $\in$  synset do
4     if lemma  $\in D$  then
5        $W \leftarrow W \cup \text{translate}(\text{synset}, D[\text{lemma}])$ 
6   for each w_synset  $\in W$  do
7     if parent(w_synset)  $\notin W$  then
8       s_synset  $\leftarrow$  parent(w_synset)
9       while s_synset  $\notin W$  and s_synset  $\neq$  top_level_synset(S)
do
10      s_synset  $\leftarrow S[\text{parent}(s\_synset)]$ 
11      parent(w_synset)  $\leftarrow$  synset
12 return W

```

Mazkur algoritmnining Input (kiritish)dan boshlab muayyan soʻz (S) qadamma-qadam PWN 3.0 versiyasidagi fayldan qidiriladi, mavjud soʻzning turkumi “Inglizcha-oʻzbekcha lugʻat”dan aniqlanadi, PWN sinsetlaridan soʻz lemmalari qidiriladi, agar lemma “Inglizcha-oʻzbekcha lugʻat”da mavjud boʻlsa, tarjima qilinadi (2-rasm).



2-rasm. uzwordnet.ldkr.org resursida *maktab* soʻziga taqdim etilgan tarjima

2-rasmdan ayon boʻlganidek, *maktab* soʻziga PWNdagi izohlari lugʻat yordamida tarjima qilingan, uning turkumi, semantik munosabatlari va sinsetlar toʻplami taqdim etilmagan. Mavjud giperonim (jins)lar qatori toʻliq emas.

UzWordNet ish jarayoni algoritmidan maʼlum boʻlganidek,

mazkur resurs uchun birginatarjimalug'atitarjimavositasi bo'lmoqda va resurs ishi to'laqonli Princeton WordNet bazasiga asoslanilgan. Shu bois mazkur resurs o'zbek tili ontologiyasi vazifasini bajara olmaydi. Ushbu xulosa UzWordNet resursidan foydalanilganda yuzaga keldi. Albatta, PWN tarkibida uning mukammallashtirilishi ko'p tilli PWN resurslarida o'zbek tilining butun imkoniyati, tabiati va xususiyatlarining namoyon bo'lishiga olib keladi. Shu bois mazkur resurs bazasi o'zbek tili tipologiyasiga muvofiq sinchkovlik bilan to'ldirilsa maqsadga muvofiq bo'ladi.

Tadqiqoti jarayonida lingvistik ontologiyani shakllantirish talablari, axborot qidiruv tezaurusini yaratish tamoyillari o'rganilildi.

Tezauruslarda bir sohadagi barcha leksik boylik iyerarxiyasi yaratilsa, lingvistik ontologiyada butun til leksikasi tarmog'i yuzaga kelishi zarur hisoblanadi. Shu bois tezauruslarni yaratish nisbatan muvaffaqiyatli va ontologiyaga nisbatan qisqa vaqtda amalga oshiriladi. So'zlar tarmog'ini noldan yaratish yoki avval mavjud bo'lganini kengaytirish esa bir necha bosqichlarni o'z ichiga oluvchi va inson mehnatidan hamda avtomatlashtirilgan tizimlardan keng foydalanishni talab qiladigan mashaqqatli jarayondir. Shu bois O'zbek tili ontologiyasi – UzNet tizimini yaratish uchun konsepsiya ishlab chiqildi.

“O'zbek tili ontologiyasini yaratish konsepsiyasi”

Tarkibi:

1. Lingvistik ontologiyani yaratishning konseptual asosi tadqiqi

2. UzNet tizimi uchun leksikografik baza yaratish

3. So'z turkumlarini teglash

4. So'zlarning shakl va semantik munosabatini belgilash

5. Sinsetlar (lug'aviy, ma'noviy va matniy sinonimlar)

Birinchi bosqich

1. Lingvistik ontologiyani yaratish metodologiyasi

1.1. Lingvistik ontologiyalarni yaratish tamoyillarini tadqiq etish.

1.2. WordNet tizimi asoslarini o'rganish.

1.3. Turkiy tillar ontologiyalari texnologiyalarini tadqiq va tahlil qilish.

1.4. UzNet tizimini yaratish loyihasini amalga oshirish.

1.5. Izohli va imlo lug'atlardagi barcha so'zlarni bazaga

kiritish.

1.6. Lingvistik ontologiya tizimi dasturiy ta'minotini ishlab chiqish.

Ikkinchi bosqich

Uznet tizimi uchun leksikografik baza

O'quv lug'atlari asosida dastlabki ma'lumotlar bazasini yaratish.

2.2. Lingvistik ontologiya uchun "O'zbek tilidagi antonimlar lug'ati"ning yangi avlodini shakllantirish va uni bazaga kiritish.

1.3. "O'zbek tilidagi ononimlar lug'ati"ning yangi avlodini shakllantirish va uni bazaga kiritish.

1.4. "O'zbek tilidagi paronimlar lug'ati"ning yangi avlodini shakllantirish va uni bazaga kiritish.

1.5. Sohalarga oid terminologik lug'atlar bazasini yaratish.

1.6. Sohalar bo'yicha tezauruslarni yaratish.

Uchinchi bosqich

3. So'z turkumlarini teglash

3.1. Izohli va imlo lug'atlardagi so'zlarning turkumlarini belgilash.

3.2. Omonim, ko'p ma'noli va polifunksional so'zlar bazasini yaratish.

3.3. So'zlarning formal izohini shakllantirish.

To'rtinchi bosqich

4. So'zlarning shakl va semantik munosabatini belgilash

4.1. Turkumlarga ajratilgan har bir so'zning ifoda va ma'no munosabatini aniqlash.

4.2. So'zlarning giponim va giperonimlarini belgilash.

4.3. Tarmoqlangan xolo-meronim munosabatini yaratish.

Beshinchi bosqich

5. Sinsetlar (lug'aviy, ma'noviy va matniy sinonimlar)

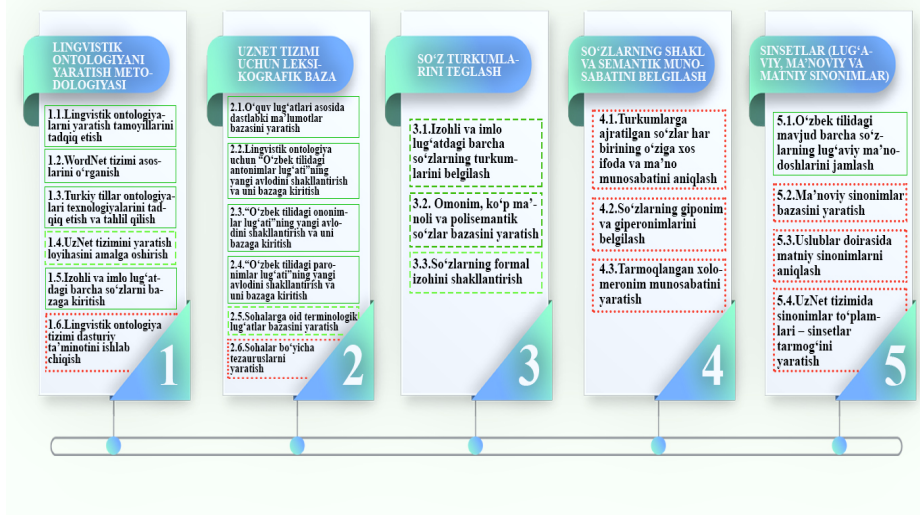
5.1. O'zbek tilida mavjud so'zlarning lug'aviy ma'nodoshlarini jamlash.

5.2. Ma'noviy sinonimlar bazasini yaratish.

5.3. Uslublar doirasida matniy (konseptual) sinonimlarni aniqlash.

5.4. UzNet tizimida sinonimlar to'plamlari – sinsetlar tarmog'ini yaratish.

“O‘ZBEK TILI ONTOLOGIYASINI YARATISH” KONSEPSIYASI



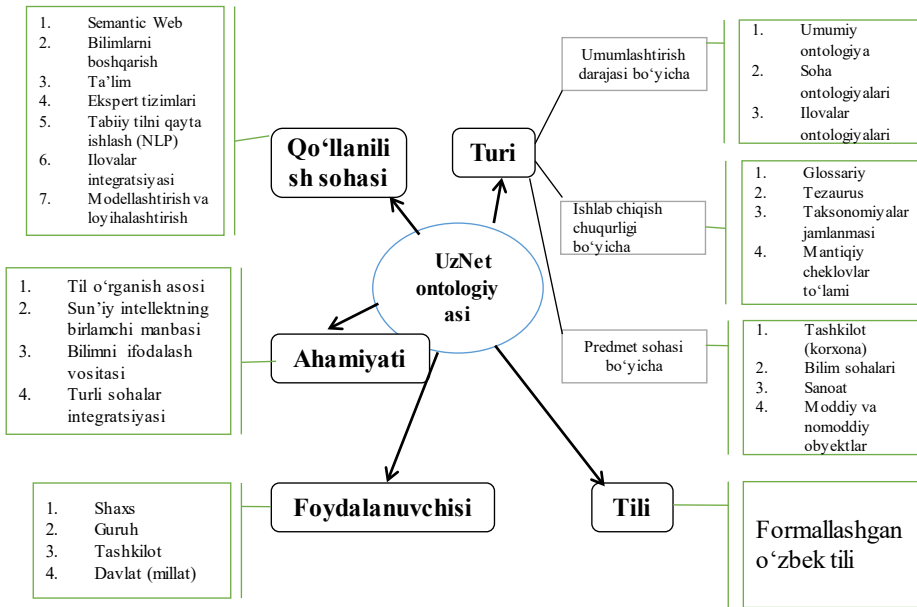
Lingvistik ontologiyani yaratish ko'p bosqichli murakkab va sinchkovlikni talab etadigan jarayon bo'lgani bois yana bir qancha tadqiqotlarni talab etadi. Yirik jamoa mehnati salmoqli natijalarni beradi.

O'zbek tili ontologiyasining asosiy tarkibiy qismlari quyidagilarni tashkil etadi:

- 1) so'z (shakl)lar / so'z birikmalari;
- 2) tushunchalar izohi;
- 3) semantik va shakliy munosabatlar (sinonimiya, giponimiya, meronimiya, antonimiya, omonimiya, paronimiya);
- 4) misollar.

Tizimda terminlar sohalararo tarmoqlanadi, sinflar tashkil etiladi.

Quyidagi chizmada O'zbek tili ontologiyasi – UzNetning qo'llanilish sohasi, tiplari, tili, ahamiyati, foydalanuvchilari tasvirlangan:



3-chizma. O'zbek tili ontologiyasi – UzNet xususiyatlari

UzNetni yaratishda ko'plab resurs va til muammolariga duch keldik. Jumladan, sinsetlarni yaratishda sinonimik munosabatlarni tizimli ravishda belgilaydigan o'zbekcha lug'at yo'qligi eng katta muammo edi. Tadqiqotda batafsil tahlil qilinganidek, lingvistik ontologiya asosida sinonimik to'plamlar turadi va sinset, ya'ni semantik munosabatlarning birlamchi elementlari ham sinonimlar sanaladi. Ayni damda mavjud "O'zbek tili sinonimlarining izohli lug'ati"da [Hojiyev] o'zbek tilidagi barcha ma'nodoshlar to'liq qamrab olinmaganligi, "O'zbek tilining ma'nodosh so'zlar o'quv lug'ati" [Otabekov, Boymatova] va "O'zbek tilining so'zlar darajalanishi o'quv lug'ati" [Bobojonov] esa faqat o'quvchilar uchun sodda va ixcham holatda tuzilganligi sababli UzNet uchun asosiy resurs vazifasini bajara olmaydi. Shu bois tadqiqot davomida ingliz, rus, turk tillari sinonim lug'atlari o'rganilib, o'zbek tili sinonimlarining zamonaviy bazasi [Abjalova va boshq] shakllantirildi.

Sinsetlar yaratishda til bilan bog'liq muammolarga duch kelindi. Jumladan, ba'zi sinsetlarda turli so'z turkumlariga mansub sinset a'zolari (*kvazisinonimlar*) mavjud. O'zbek tilida bir xil ma'noga ega bo'lgan shunday so'zlar borki, ularning har biri turli so'z turkumlariga mansub hisoblanadi. Masalan, *globalizatsiya*, *lemmatizatsiya* – jarayon nomi, ot turkumidagi so'zlar va *globalashuv*, *lemmalash* – harakat nomi, fe'l turkumidagi so'zlar sanaladi. Bunday holat sinonimlarning morfologik jihatdan noto'g'ri guruhlanishiga olib kelishi mumkin: kvazisinonimlarni avtomatik ravishda

sinonimik qator mansub bo'lgan so'z turkumiga kiritib bo'lmaydi, balki mutaxassis tomonidan bittalab turkumlanishi va shu asosda guruhlanishi talab qilinadi. Bu borada KeNet resursini yaratish tajribasiga tayanildi. KeNetni o'rganish natijasida giperonimik munosabatlarni o'rnatishda ba'zi PWN sinsetlarining KeNet sinsetlariga mos kelmasligi aniqlangan edi. Bu muammoni yechish uchun KeNet loyihasi jamoasi iyerarxiyaning yuqori qismlaridagi shunday sinsetlardan ikki til uchun muvofiq, iyerarxiya daraxtining quyi qismlaridagi sinsetlarni yuqori qismlari bilan bog'laydigan yangi sinsetlar yaratilgan, aks holda iyerarxiyada ba'zi o'tish tugunlari yetishmagan bo'lar edi.

Tadqiqot jarayonida UzNet ontologiyasini yaratishda WordNetga butunlay tayana olmaslik va undagi boshqa resurslarga murojaat qila olmaslik ma'lum bo'ldi. Sababi WordNet resurslarini yaratishda ko'p bosqichlarning avtomatik bajarilishi katta muammolarni yuzaga keltiradi. Chunki bir tabiiy til (ingliz tili) xususiyatlari boshqa bir tabiiy til (jumladan, o'zbek tili) xususiyatlari bilan mos emas. Shuning uchun UzNetni yaratishda dastlabki bosqichlarni qo'lda bajarilishga to'g'ri keladi. Bu jarayonda qoidalarga asoslangan metodga tayaniladi, aksariyat ma'lumotlar qayta ishlangandan so'ng stoxastik metod asosida baza boyitiladi.

Tabiiy til tushunchalari, so'zlar va ifodalar orasidagi munosabatlarni shakllantirishda tilning ijtimoiy tabiati bilan bog'liq muammolar yuzagakeladi. Bundantashqari, ontologiyatushunchalari o'rtaidagi munosabatlarni aniq va muqim holda o'rnatish ham tilning pragmatik jihati bilan bog'liq qiyin masala hisoblanadi, ayniqsa, "leksik birlik – tushuncha (yoxud uning izohi)" aloqalarini tashkil qilishda ham muammolar yuzaga keladi. Bu masalada, tabiiy tildagi har bir leksik birlik (LB)ning to'liq va mukammal izohlari shakllantirilishi talab qilinadi. Buning uchun, avvalo, (1) butun so'z boyligi jamlanishi, (2) har bir LBning turkumi aniqlanishi va (3) o'sha LBning aniq hamda ixcham izoh(lar)i darajalangan (yoxud asosiy tushunchani anglatishi jihatidan tartibli raqamlangan) tarzda shakllantirilishi darkor. O'zbek tili ontologiyasini yaratishda ilgari surilgan **keng qamrovlilik, formal til aspekti, muvofiqlik va semantik simmetriya** tamoyillarida barcha tillar uchun umumiylik kasb etuvchi ushbu zarur qadamlar e'tiborga olingan.

O'zbek tilida fors-tojik tilidan o'zlashgan *osmon* so'zining *ko'k*, *falak* [a], *gardun* [f-t], *samo* [a] kabi sinonimlari mavjud. Lingvistik ontologiyada **sinsetlar** quyidagicha darajalanadi:

Joylashuv o'рни	Sinsetlar	Izohi
Yuqori qator sinsetlar	<i>osmon, falak, gardun, samo</i>	Yer ustida gumbaz shaklida ko'rinib turadigan havo qatlami
Quyi qator sinsetlar	<i>ko'k</i>	Yer ustida gumbaz shaklida ko'rinib turadigan havo qatlami
Ekzemplarlar	<i>Osmon</i>	ism
	<i>Osmon</i>	"Ummon" guruhi tomonidan ijro etilgan qo'shiq nomi

Bo'yoqdorligiga ko'ra darajalanishi: *osmon – ko'k – samo – falak – gardun;*

Giperonimi: *atmosfera* (Yerni va boshqa ba'zi planetalarni o'rab turgan gazsimon qobiq);

Giponimi: *ko'k osmon, ko'k;*

Xolonimi (qisman): *Yer, dunyo, olam, globus* (quyosh sistemasining 3-yo'ldoshi);

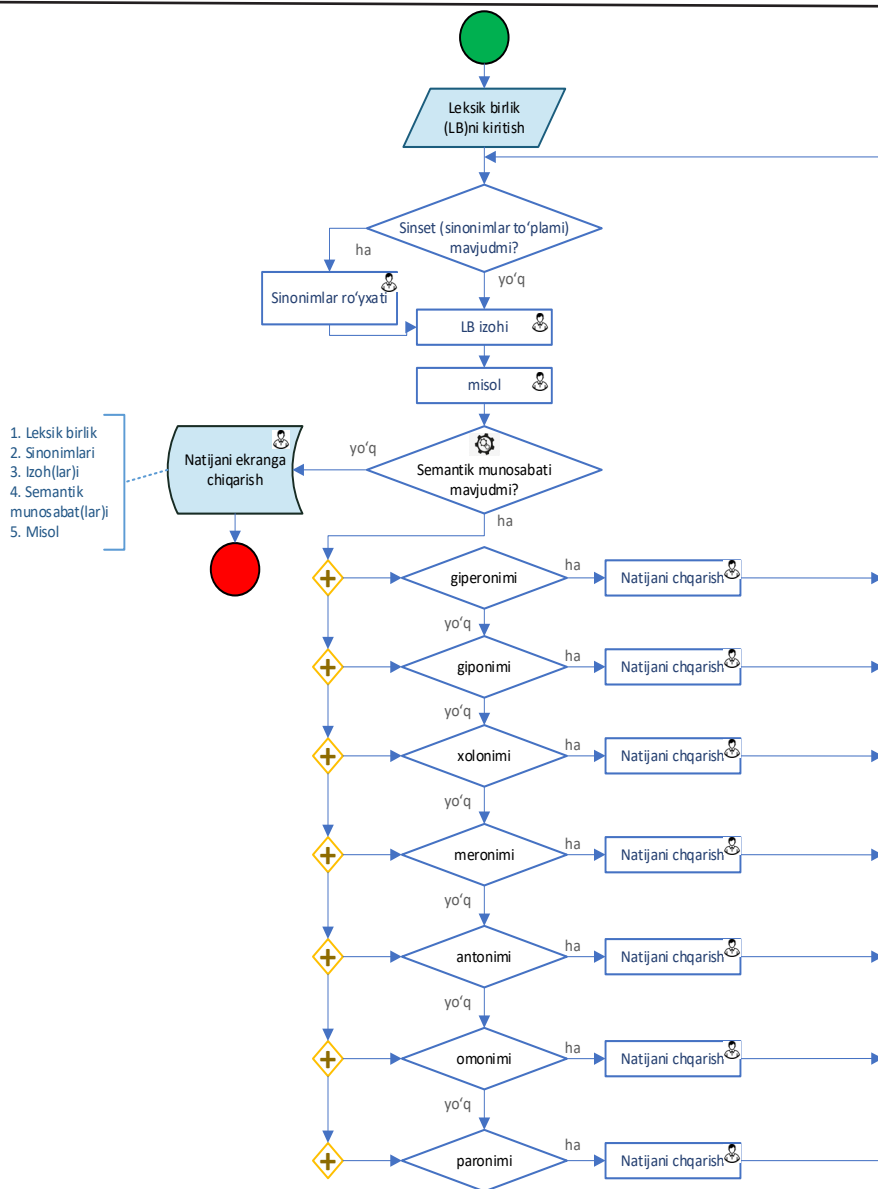
Meronimi: *bulut* (suv/muz bo'laklarining balandlikda ko'rinadigan massasi); *kamalak* (Quyosh nurining yomg'ir, favvora tomchilarida sinishi natijasida havoda turli rangda yarim doira shaklida ko'rinadigan yo'l);

Antonim(lar)i: yer, zamin, arz [a].

Ma'lum bo'lganidek, *osmon* so'zi o'ziga *ko'k* rangni va antonimi hisoblanmish *yer* so'zini bog'laganligi uchun *osmon* orqali *ko'k* va *yer* sinsetlariga o'tish mumkin.

Tizimning biznes-jarayon modeli esa graflar nazariyasidagi rekursiv algoritmlar asosida quyidagicha bo'ladi (4-chizma).

UzNetni yaratishda duch kelingan eng katta muammo cheklangan resurslar bilan ishlash bo'lgan. Chunki bu sohadagi tadqiqotlarning aksariyati ingliz tili kabi yaxshi o'rganilgan (formal) tillarga qaratilgan. Turk tilining turli morfologik va sintaktik xususiyatlari ham ushbu tadqiqotlarda qo'llanilgan aniq yondashuvlarni qabul qilishga to'sqinlik qildi. Tillarning struktur tuzilishidagi farqlar, asosan, morfologik tahlilda muammo tug'dirdi, shuningdek, ingliz tilidagi semantik munosabatlarni to'g'ridan-to'g'ri o'zbek tiliga ko'chirib bo'lmaydigan, ingliz va o'zbek tillaridagi bir-biriga muvofiq kelmaydigan holatlarga duch kelindi. Ushbu nomuvofiqlik o'zbek tili uchun ham ichki, ham tillararo tashqi semantik munosabatlarda yuzaga keldi.



4-chizma. O'zbek tili ontologiyasi biznes-protsess modeli

MikroKosmos yoki CYC kabi ko'plab munosabatlarga ontologiyalardan foydalanish muammosi ikki jihat bilan bog'liq. Birinchidan, yangi soha bo'yicha bunday semantik resurs yaratish nihoyatda qiyin, katta mablag' va ko'p vaqtni talab etadi. Ikkinchidan, bunday tizimlardagi ko'p sonli munosabatlar matnni qayta ishlashda ham chalg'ituvchi omil bo'lib xizmat qilishi mumkin, chunki tavsiflangan munosabatlarning faqat bir qismi muayyan bir kontekstda qo'llanilishi mumkin, qolgan munosabatlar esa keraksiz bo'ladi yoki noto'g'ri xulosalarga olib kelishi mumkin. Shu bois

ontologiyani yaratishda fidoyi, ziyrak va zukko tilshunoslar talab etiladi, ulardan esa, o'z navbatida, pragmatik tahlil olish malakasining ham bo'lishi talab qilinadi. Zero, kontekst bo'yicha munosabatlarning qo'llanilishini avtomatik ravishda baholash o'ta qiyin.

WordNet tipidagi lingvistik resurlar tilshunoslik an'alariga muvofiq tilning lug'at boyligini tavsiflash uchun yaratilgan. Biroq har qanday axborot tizimi nafaqat umumiy lug'at, balki muayyan fan sohasi va uning terminologiyasi bilan ham shug'ullanadi. WordNet asosida terminologik resurslarni yaratishga urinishlar tahlil qilinganda xulosa shunday bo'ldiki, WordNet strukturasi terminologiyalarni tavsiflash uchun moslashtirilmagan. So'z turkumlarining alohida tavsifi, bir-biriga bog'liq bo'lmagan ma'nolarning juda katta to'plami, ko'p tarkibli ifodalarni kiritish tamoyillarining yetarli darajada ishlab chiqilmaganligi – bularning barchasi WordNet modeli asosida yaratilgan terminologik resurslarni ishlab chiqish va ulardan foydalanishda muammolarni yuzaga keltiradi.

Xulosa

Xulosa qilib aytganda, lingvistik ontologiyalar mashina tarjimasi, savol-javob tizimlari, ma'lumot qidirish, bilimlarni olish tizimlari, kompyuter va shaxs o'rtasidagi muloqotni o'tkazish tizimlari, tilni tushunish tizimlari, shuningdek, bilimlarni namoyish qilish, sun'iy intellekt va kompyuter ma'lumotlarini qayta ishlash bilan bog'liq ko'plab muammolarni hal qilishda muhim ahamiyatga ega. Xususan, tilshunoslik sohasida ontologiyalardan matn korpusining semantik Abstraktsi, mashina tarjimasi, ko'p ma'nolilikni avtomatik hal qilish va kontekstga asoslangan omonimiyani aniqlash, quyi darajadagi ontologik turdagi resurs, lug'at va tezauruslarni yaratishda foydalaniladi. Bundan tashqari, ko'p tilli ontologiyalardan tarjimonlar tomonidan bilimlar va tegishli lug'atni o'z ichiga olgan ma'lumot manbalari sifatida foydalaniladi. Shu bois UzNet o'zbek tili ontologiyasini yaratish muhim ahamiyat kasb etadi.

Adabiyotlar

- Agostini, A., Usmanov, T. and other. 2021. "UZWORDNET: A Lexical- Semantic Database for the Uzbek Language. In S. Bosch, C. Fellbaum, M. Griesel, A. Rademaker and P. Vossen, editors, Proceedings of the Eleventh International Global Wordnet Conference. P. 8.
- Abjalova, M., Rashidova, U., Rahimov, S. 2019. O'zbek tilidagi sinonim so'zlarning ma'lumotlar bazasi. Guvohnoma № BGU 00380. Toshken. (mualliflik guvohnomasi).
- Bobojonov, Sh., Islomov, I. 2007. O'zbek tilining so'zlar darajalanishi o'quv

- lug'ati: Maktab o'quvchilari uchun. – Toshkent: Yangi asr avlodi. – 38 b.
- Eiji, Aramaki, Takeshi, Imai, Masayo, Kashiwagi, Masayuki, Kajino, Kengo, Miyo and Kazuhiko Ohe. 2005. Toward medical ontology using Natural Language Processing. //URL: <http://www.m.u-tokyo.ac.jp/medinfo/ont/paper/2005-aramaki-1.pdf>.
- Hojiyev, A. (1974). O'zbek tili sinonimlarining izohli lug'ati. – Toshkent: O'qituvchi. 308 b.
- Otabekov, Sh., Boymatova, B. 2007. O'zbek tilining ma'nodosh so'zlar o'quv lug'ati: Maktab o'quvchilari uchun. – Toshkent: Yangi asr avlodi. – 49 b.
- Гаврилова, Т.А. Использование онтологий в системах управления знаниями // http://big.spb.ru/publications/bigspb/km/use_ontology_in_suz.shtml
- Гладун, А.Я., Рогушина, Ю.В. 2006. Онтологии в корпоративных системах. Часть II // Корпоративные системы №1. //URL: <http://www.management.com.ua/ims/ims116.html>
- Клещев, А.С., Шалфеева, Е.А. 2005. Классификация свойств онтологий. Онтологии и их классификации. – Владивосток: ИАПУ ДВО РАН. – С. 3.
- Росеева, О. И., Загорюлько, Ю.А. 2001. Организация эффективного поиска на основе онтологий // Труды Международного семинара Диалог>2001 по компьютерной лингвистике и ее приложениям, т.2. – [Электронный ресурс]: [Статья]. – Режим доступа: <http://www.dialog-21.ru/materials/archive.asp?id=7029&y=2001&vol=6078>
- <http://compling.hss.ntu.edu.sg/omw/>
- <http://globalwordnet.org/>
- <http://globalwordnet.org/resources/wordnets-in-the-world/>
- <https://github.com/LDKR-Group/UzWordnet>
- <https://uzwordnet.ldkr.org/>

The concept of designing the creation of uznet linguistic ontology

Manzura Abjalova¹

Abstract

It should be pointed out that in order to improve modern branches of human activity which requires knowledge it is important to increase of the role of computer technologies worldwide. At present, because the information flow is rising considerably, now there is necessity of protection, officialization and arrangement of information in a sequence and seeking new ways of processing it automatically as well. Due to this, there is an increasing interest in a wide range of information bases which can be used in practical terms. Especially, the demand is really huge for systems based on neural-nets that takes out any kind of information from the text without human interaction. The half of 20th century saw the emergence of the semantic webs together with global branch and it has been provided with extra tags carrying information about element semantics of hypertext pages. Inseparable part of semantic web is the concept of ontology which is lexical database consisting of a branch of words.

This article covers the creation of the Uzbek language ontology – the concept of the system of UzNet, the basics of its design, the business process model (BPMN) of UzNet ontology.

Key words: *Linguistic ontology, semantic web, Uznet, semantic relationship, concept, designing.*

Reference:

- Agostini, A., Usmanov, T. and other. "UZWORDNET: A Lexical- Semantic Database for the Uzbek Language. In S. Bosch, C. Fellbaum, M. Griesel, A. Rademaker and P. Vossen, editors, Proceedings of the Eleventh International Global Wordnet Conference (GWC-2021), – P. 8.
- Abjalova, M., Rashidova, U., Rakhimov, S. 2019. Database of synonyms of the Uzbek language. Certificate № BGU 00380. – Tashkent, (author's license).

¹Abjalova Manzura Abdurashetovna – Doctor of Philosophy in Philology (PhD), Associate Professor of Computer Linguistics and Digital Technologies of Tashkent State University of Uzbek Language and Literature named after Alisher Navoi.

E-mail: abjalovamanzura@navoiy-uni.uz

ORCID: 0000-0002-1927-2669

For reference: Abjalova, M.A., Gulomova N.S. 2022. "The Concept of Designing the Creation of UzNet Linguistic Ontology". *Uzbekistan: language and culture. Applied Philology* 1(5): 123-139.

- Bobojonov, Sh., Islomov, I. 2007. Educational Dictionary of the level of vocabulary of the Uzbek language: for schoolchildren. Tashkent: Yangi asr avlodi, – 38 p.
- Eiji, Aramaki, Takeshi Imai, Masayo Kashiwagi, Masayuki Kajino, Kengo Miyo and Kazuhiko Ohe. Toward medical ontology using Natural Language Processing. //URL: <http://www.m.u-tokyo.ac.jp/medinfo/ont/paper/2005-aramaki-1.pdf>.
- Khojiev, A. 1974. Explanatory Dictionary of synonyms of the Uzbek language. Tashkent: O'qituvchi, 308 p.
- Otabekov, Sh., Boymatova, B. 2007. Educational Dictionary of meaningful words of the Uzbek language: for schoolchildren. – Tashkent: Yangi asr avlodi, 49 p.
- Gavrilova, T.A. 2006. The use of ontologies in knowledge management systems // http://big.spb.ru/publications/bigspb/km/use_ontology_in_suz.shtml.
- Gladun, A.Ya., Rogushina, Yu.V. Ontologies in corporate systems. Part II // Corporate systems №1. //URL: <http://www.management.com.ua/ims/ims116.html>.
- Kleshchev, A.S., Shalfeeva, E.A. 2005. Classification of ontology properties. Ontologies and their classifications. – Vladivostok: IAPU FEB RAS, – P. 3.
- Rosseeva, O. I., Zagorulko, Yu.A. 2001. Organization of effective search based on ontologies // Proceedings of the International Seminar Dialog'2001 on Computational Linguistics and its applications, vol.2, [Electronic resource]: [Article]. Access mode: <http://www.dialog-21.ru/materials/archive.asp?id=7029&y=2001&vol=6078>.
- <http://compling.hss.ntu.edu.sg/omw/>
- <http://globalwordnet.org/>
- <http://globalwordnet.org/resources/wordnets-in-the-world/>
- <https://github.com/LDKR-Group/UzWordnet>
- <https://uzwordnet.ldkr.org/>