



KORPUS LINGVISTIKASIDA BILIMLAR BAZASI VA ONTOLOGIK MODELLASHTIRISH

Mamarajabov Abduvali Yuldosh o'g'li
TerDU O'zbek tili va adabiyoti kafedrasida o'qituvchisi

Annotatsiya. Ushbu maqolada korpus lingvistikasida bilimlar bazasini yaratish va ontologik modellashtirish masalalari o'rganiladi. Ontologik modellar til ma'lumotlarini formal tizim sifatida ifodalash imkonini beradi. Bilimlar bazasi esa ushbu ma'lumotlarni saqlash va qayta ishlash uchun asos bo'lib xizmat qiladi. Tadqiqotda lingvistik resurslarni integratsiyalash, semantik tarmoqlar va avtomatik tizimlar bilan bog'liqlik masalalari tahlil qilinadi.

Kalit so'zlar: bilimlar bazasi, ontologik model, korpus lingvistikasi, semantik tarmoq, ma'lumotlarni integratsiya, sun'iy intellekt, lingvistik resurs, ma'lumotlar modeli, avtomatik tizim, konseptual struktura.

Kirish

Korpus lingvistikasi zamonaviy tilshunoslikda katta hajmdagi matnlarni tizimli tahlil qilish imkonini beruvchi muhim yo'nalishdir. Bunday tizimlarda faqat matnni yig'ish emas, balki uni bilimlar bazasi sifatida tashkil etish ham dolzarb ahamiyatga ega. Ontologik modellashtirish lingvistik ma'lumotlarni formal va mantiqiy tizimga keltirish imkonini beradi. Bu esa sun'iy intellekt, mashina o'rganishi va avtomatik tahlil tizimlarining rivojlanishiga xizmat qiladi. Mazkur maqolada korpus lingvistikasida bilimlar bazasi va ontologik modellashtirishning nazariy asoslari va amaliy qo'llanilishi yoritiladi.

Asosiy qism

Korpus lingvistikasida bilimlar bazasi va ontologik modellashtirish zamonaviy tilshunoslikning eng muhim yo'nalishlaridan biri bo'lib, u tilni nafaqat kommunikativ vosita, balki formal bilimlar tizimi sifatida o'rganishga imkon beradi. An'anaviy korpus lingvistikasida asosiy e'tibor matnlarni yig'ish, ularni morfologik va sintaktik jihatdan tahlil qilishga qaratilgan bo'lsa, zamonaviy yondashuvlarda korpus ma'lumotlarini bilimlar bazasi sifatida tashkil etish ustuvor ahamiyat kasb etmoqda. Bu esa korpusni oddiy matnlar to'plamidan kompleks semantik tizimga aylantiradi. Bilimlar bazasi lingvistik korpusda til birliklari, ularning ma'nolari va o'zaro munosabatlarini tizimli ravishda saqlash va qayta ishlash imkonini beruvchi formal tuzilma sifatida qaraladi. Bunday bazada har bir lingvistik birlik mustaqil element



emas, balki ma'lum bir konseptual tizimning tarkibiy qismi sifatida ifodalanadi. Shu sababli bilimlar bazasi korpus lingvistikasida semantik integratsiyani ta'minlovchi asosiy infratuzilma hisoblanadi [3, 11].

Ontologik modellashtirish esa ushbu bilimlar bazasini formal va ierarxik tizimga keltirish jarayonidir. Ontologiya yordamida til birliklari o'rtasidagi mantiqiy va semantik munosabatlar aniqlanadi va ular "is-a", "part-of", "causes", "related-to" kabi formal bog'lanishlar orqali ifodalanadi. Bu yondashuv korpusdagi ma'lumotlarni faqat saqlash emas, balki ularni tushunish va tahlil qilish imkonini ham beradi. Natijada korpus lingvistikasi bilimlar muhandisligi bilan integratsiyalashgan yangi bosqichga chiqadi.

C. M. Keet tomonidan olib borilgan tadqiqotlarda ontologik modellashtirishning asosiy vazifasi real dunyo bilimlarini formal semantik struktura sifatida ifodalash ekanligi ta'kidlanadi. Uning fikricha, ontologiya nafaqat ma'lumotlarni tasvirlash vositasi, balki bilimlarni tushunish va qayta ishlash mexanizmi hamdir [4, 18]. Lingvistik korpuslar kontekstida bu yondashuv til birliklarini konseptual obyektlarga aylantirish imkonini beradi.

Bilimlar bazasi va ontologik modelning integratsiyasi natijasida korpuslar grafik ma'lumotlar tizimiga aylanadi. Bu grafik modelda har bir tugun (node) ma'lum bir lingvistik konseptni ifodalaydi, qirralar (edges) esa ularning o'zaro munosabatini ko'rsatadi. Bunday yondashuv knowledge graph (bilimlar grafigi) yaratish imkonini beradi. Knowledge graphlar zamonaviy sun'iy intellekt tizimlarida keng qo'llanilib, ular orqali murakkab semantik so'rovlarni qayta ishlash, avtomatik xulosa chiqarish va tavsiya tizimlarini yaratish mumkin.

Ontologik modellashtirish jarayoni bir nechta bosqichdan iborat bo'lib, avvalo lingvistik birliklar aniqlanadi, so'ng ular konseptual kategoriyalarga ajratiladi va nihoyat ular o'rtasidagi munosabatlar formal ravishda belgilanadi. Bu jarayon yarim avtomatik yoki avtomatik tarzda amalga oshirilishi mumkin. Avtomatik yondashuvlarda mashina o'rganishi algoritmlari muhim rol o'ynaydi, chunki ular katta hajmdagi korpus ma'lumotlarini tahlil qilib, konseptual bog'lanishlarni aniqlashga yordam beradi.

Abduraxmonova N. tomonidan o'zbek elektron korpusi ustida olib borilgan tadqiqotlarda ham lingvistik ma'lumotlarni tizimlashtirishda bilimlar bazasining ahamiyati alohida ta'kidlangan. Uning modelida har bir til birligi faqat leksik birlik sifatida emas, balki semantik va funksional jihatdan bog'langan tizim elementi sifatida qaraladi [1, 9]. Bu yondashuv o'zbek tilini raqamli formatda tizimli o'rganish imkonini kengaytiradi.



Ontologik modellashtirishning muhim jihatlaridan biri uning universal xarakterga ega ekanligidir. Ya'ni, yaratilgan ontologik model turli tillarga moslashtirilishi mumkin. Bu esa multilingual korpuslar yaratishda katta ahamiyatga ega. Masalan, ingliz, rus va o'zbek tillaridagi bir xil konseptlar yagona ontologik tizimda birlashtirilishi mumkin. Bu esa tarjima tizimlari va cross-lingual NLP modellarini rivojlantirish uchun asos bo'lib xizmat qiladi. Shuningdek, bilimlar bazasi va ontologiya integratsiyasi axborot qidiruv tizimlarining aniqligini sezilarli darajada oshiradi. An'anaviy qidiruv tizimlari faqat kalit so'zlarga asoslangan bo'lsa, ontologik tizimlar ma'no asosida ishlaydi. Bu esa foydalanuvchi so'rovining kontekstini hisobga olish imkonini beradi. Natijada semantik qidiruv tizimlari ancha aniq va samarali natijalar beradi.

Zamonaviy NLP tizimlarida, ayniqsa neyron tarmoqlarga asoslangan modellar (BERT, GPT) keng qo'llanilayotgan bo'lsa-da, ular faqat statistik bog'lanishlarga tayanadi. Ontologik modellar esa bu statistikani konseptual bilimlar bilan boyitadi va gibrid tizimlar yaratish imkonini beradi. Bu esa tilni qayta ishlashda yanada chuqur semantik tushunishni ta'minlaydi.

Biroq ontologik modellashtirish jarayonida bir qator muammolar ham mavjud. Eng asosiy muammolardan biri – lingvistik bilimlarning standartlashtirilmaganligidir. Turli tadqiqotchilar tomonidan yaratilgan ontologiyalar bir-biri bilan to'liq mos kelmasligi mumkin. Bundan tashqari, katta hajmdagi korpuslar uchun to'liq avtomatik ontologiya yaratish hali ham murakkab masala bo'lib qolmoqda.

Shunga qaramay, bilimlar bazasi va ontologik modellashtirish lingvistik korpuslarning kelajakdagi rivojlanish yo'nalishini belgilab bermoqda. Bu yondashuv tilni nafaqat kommunikativ tizim sifatida, balki formal bilimlar tizimi sifatida o'rganish imkonini beradi. Natijada tilshunoslik informatika, sun'iy intellekt va bilimlar muhandisligi bilan chuqur integratsiyalashgan interdisiplinar sohaga aylanmoqda.

Korpus lingvistikasida bilimlar bazasi va ontologik modellashtirishning rivojlanishi raqamli gumanitar fanlar paradigmasi bilan bevosita bog'liq bo'lib, u tilni strukturaviy tizim sifatida emas, balki dinamik bilimlar ekotizimi sifatida talqin qilishga olib kelmoqda. Bunday yondashuvda korpus faqat matnlar yig'indisi emas, balki semantik jihatdan bog'langan konseptlar tarmog'i sifatida qaraladi. Bu esa til ma'lumotlarini qayta ishlashda yangi metodologik asosni shakllantiradi, ya'ni ma'lumotlar ombori emas, balki bilimlar modeli yaratiladi.

Bilimlar bazasining asosiy xususiyati shundaki, u lingvistik birliklarni izolyatsiyalangan element sifatida emas, balki o'zaro bog'langan tizim elementlari sifatida saqlaydi. Har bir birlikning semantik atributlari, sintaktik funksiyalari va pragmatik kontekstlari birgalikda ifodalanadi. Shu sababli bilimlar bazasi korpus



lingvistikasida integrativ platforma vazifasini bajaradi. Ushbu platforma orqali turli darajadagi lingvistik ma'lumotlar yagona strukturaga birlashtiriladi va bu ma'lumotlar keyinchalik avtomatik tahlil tizimlari uchun asos bo'lib xizmat qiladi [3, 14].

Ontologik modellashtirish bu jarayonni yanada formal va aniq bosqichga olib chiqadi. Ontologiya yordamida bilimlar bazasidagi ma'lumotlar mantiqiy qoidalar asosida tartibga solinadi. Bu jarayonda asosiy e'tibor konseptlar o'rtasidagi ierarxik va semantik munosabatlarga qaratiladi. Masalan, "o'qituvchi" konsepti "inson" kategoriyasining bir qismi sifatida belgilanadi va u "ta'lim jarayoni", "bilim uzatish" kabi konseptlar bilan bog'lanadi. Bu esa til birliklarini nafaqat tasvirlash, balki ularni tushuntirish imkonini ham beradi.

C. M. Keetning ilmiy qarashlariga ko'ra, ontologik modellar bilimlarni formal tizimga keltirish orqali ularni kompyuter tizimlari tomonidan qayta ishlashga moslashtiradi. Bu jarayon knowledge representation (bilimlarni ifodalash) muammosini hal qilishda muhim rol o'ynaydi. Ontologiya yordamida bilimlar faqat saqlanmaydi, balki ular o'rtasidagi mantiqiy xulosalar ham avtomatik ravishda chiqarilishi mumkin [4, 27]. Bilimlar bazasi va ontologik modellashtirish integratsiyasi natijasida korpus lingvistikasi yangi bosqichga — semantik hisoblash (semantic computing) bosqichiga o'tadi. Bu bosqichda til birliklari ustida faqat tahlil emas, balki hisoblash operatsiyalari ham amalga oshiriladi. Masalan, ma'lum bir konseptga tegishli barcha lingvistik birliklarni avtomatik aniqlash, ularning o'zaro bog'lanishlarini vizuallashtirish va semantik tarmoqlarni qurish mumkin bo'ladi.

Shu bilan birga, ontologik modellar lingvistik resurslarni standartlashtirishda ham muhim rol o'ynaydi. Turli korpuslar va lingvistik bazalar o'rtasida moslikni ta'minlash uchun yagona ontologik strukturaga ehtiyoj mavjud. Bu esa xalqaro lingvistik resurslarni integratsiya qilish imkonini beradi. Natijada multilingual korpuslar yaratish, ularni solishtirish va umumiy semantik model qurish osonlashadi.

N. Abduraxmonova tomonidan olib borilgan tadqiqotlarda o'zbek tilining elektron korpusini yaratishda ontologik yondashuvning muhimligi ta'kidlangan. Uning modelida til birliklari faqat morfologik yoki sintaktik birlik sifatida emas, balki semantik tarmoq elementlari sifatida qaraladi. Bu esa o'zbek tilining raqamli resurslarini global NLP tizimlariga integratsiya qilish imkonini beradi [1, 10].

Bilimlar bazasining yana bir muhim jihati uning dinamikligidir. Til doimiy ravishda o'zgarib boruvchi tizim bo'lgani uchun bilimlar bazasi ham statik emas, balki dinamik ravishda yangilanib turishi kerak. Ontologik modellar bu jarayonni qo'llab-quvvatlaydi, chunki ular yangi konseptlarni qo'shish va mavjud munosabatlarni qayta tashkil qilish imkonini beradi. Bu esa korpus lingvistikasini zamonaviy til



evolyutsiyasiga moslashtiradi. Shuningdek, ontologik modellashtirish sun'iy intellekt tizimlarida avtomatik qaror qabul qilish jarayonlarini qo'llab-quvvatlaydi. Masalan, semantik qidiruv tizimlarida foydalanuvchi so'roviga mos natijalarni faqat kalit so'zlar asosida emas, balki konseptual bog'lanishlar asosida ham aniqlash mumkin bo'ladi. Bu esa axborot qidiruv samaradorligini sezilarli darajada oshiradi. Biroq ontologik modellashtirishning murakkabligi uning ko'p resurs talab qiluvchi jarayon ekanligida ham namoyon bo'ladi. Katta hajmdagi korpuslar uchun to'liq ontologiya yaratish ko'p vaqt va yuqori darajadagi lingvistik bilimlarni talab qiladi. Bundan tashqari, avtomatik tizimlarda semantik xatoliklar ehtimoli mavjud bo'lib, bu inson ekspertizasi bilan tekshirish zaruratini keltirib chiqaradi.

Xulosa

Korpus lingvistikasida bilimlar bazasi va ontologik modellashtirish tilni formal va semantik jihatdan chuqur tahlil qilish imkonini beradi. Ushbu yondashuv natijasida korpuslar oddiy matnlar to'plamidan bilimlar grafigiga aylanadi. Ontologiya lingvistik birliklar o'rtasidagi mantiqiy bog'lanishlarni tizimlashtirib, sun'iy intellekt tizimlarining aniqligini oshiradi. Umuman olganda, bu yondashuv zamonaviy tilshunoslikni raqamli va interdisiplinar ilmiy yo'nalish sifatida rivojlantirishda muhim metodologik asos bo'lib xizmat qiladi.

Foydalanilgan adabiyotlar ro'yxati

1. Abduraxmonova N. *Elektron korpus va lingvistik modellashtirish* [Elektron resurs]. – Zenodo, 2021. – URL: <https://doi.org/10.5281/zenodo.5528953> (murojaat sanasi: 09.05.2026).
2. Abjalova M. *Korpus lingvistikasi va bilimlar bazasi masalalari* [Elektron resurs]. – CTCL.uz, 2023. – URL: <https://ctcl.uz/index.php/uzmu/article/view/108>
3. *Ontologik bilimlar bazasini yaratish muammolari* [Elektron resurs]. – MyScience.uz, 2023. – URL: <https://myscience.uz/index.php/linguistics/en/article/view/104>
4. Keet C. M. *Ontology learning from corpora* [Elektron resurs]. – arXiv, 2017. – URL: <https://arxiv.org/abs/1709.01991>
5. Noy N. F. *Ontology development 101* [Elektron resurs]. – Stanford University, 2001. – URL: https://protege.stanford.edu/publications/ontology_development/ontology101.pdf