

MASHINA TARJIMASI UCHUN OCHIQ KORPUSLI PARALLEL KORPUS BAZALARI YARATISH

Olimjonov Axrorjon Farxodjon o'g'li

Toshkent axborot texnologiyalari universiteti,

Axborot va multimedia texnologiyalari yo'nalishi 2-kurs magistranti

E-mail: axrorbekjon98@gmail.com

Abidova Shaxnoza Baxodirovna

Ilmiy rahbar:

ANNOTATSIYA

Ushbu maqolada Mashina tarjimasi uchun ochiq korpusli parallel korpus bazalari yaratish hamda o'zbek tilining fan va sohalar bo'yicha ochiq manbali terminologik ma'lumotlar bazasini loyihalashtirish va yaratish masalasi tahlilga tortilgan. Unga ko'ra ochiq manbali terminologik baza doimiy ravishda to'ldirilib borilishi hamda fan va sohalar bo'yicha yangi kirib kelgan lingvistik terminlarni bazaga kiritish imkoniyati mavjudligi bilan ahamiyatlidir.

Kalit so'zlar: terminologik baza, lingvistik ta'minot, qidiruv tizimi, mashina tarjimasi, o'zbek tili.

KIRISH

Globalashuv jarayoni nafaqat insonlarning turmush tarzi va madaniyati, balki kishilar muloqotining asosiy vositasi bo'lgan tilga ham o'z ta'sirini ko'rsatmoqda. Asrimizning rivojlanishi natijasida yangi sohalarining yuzaga kelishi va kundan- kunga yangi terminlarning paydo bo'lishiga olib kelmoqda.

Matnlarni avtomatik tarjima qilishda ilmiy va texnikaviy terminlarni o'rganish va ularning muqobil variantlarini tarjima qilishda terminografiya¹ (terminologik lug'atshunoslik) muhim masala hisoblanadi. "Ko'p tilli terminografiyaning dolzarbligi tez fursatda o'sib borayotgan avtomatlashgan lug'atlar va terminologik ma'lumotlar bankining paydo bo'lgani bilan izohlanadi [2]". Ayrim statistik ma'lumotlarga qaraganda XX asrda Germaniyada barcha ilmiy-texnikaviy terminlar 1 mln.ni tashkil etgan. XX asr o'rtalariga kelib elektrotexnikaga terminlar miqdori 4 mln.ga yetgan.

Terminologik lug'atlarni strukturalash va ularni ma'lumotlar bazasida loyihalash ahamiyatga ega. Mashina tarjimasida terminografiya alohida o'rganilishi uning lingvistik ta'minotda asosiy qism ekanligini ko'rsatadi. Mashina tarjimasida

terminlarning alohida sohaviy lug'atlari yaratilishi, tarjimaning sifatini oshiradi. Terminologiya hamisha o'zgarib boruvchi soha bo'lgani uchun rivojlangan davlatlarda bu yo'nalish jiddiy tadqiq etilgan. Ammo o'zbek tilshunosligida kompyuter terminografiyasi deyarli o'rganilmagan. Terminografiyada quyidagi masalalar o'rganiladi:

- 1) u yoki bu terminning ikkinchi bir tilda aynan mavjud bo'lmashligi;
- 2) terminlarning so'z birikmalari tarzida ifodalanishi;
- 3) matnning uslubini aniqlash va ilmiy terminlarning ma'nolarini uslubga ko'ra o'zgarishi;
- 4) muayyan soha doirasida bir leksemaning turli ma'noda qo'llanilishi.
- 5) muayyan terminning bir necha sohada turli shaklda ifodalanishi

Masalan, texnika sohasida qo'llaniladigan inglizcha *pin* termini ot so'z turkumi sifatida quyidagi ma'nolarga ega: *barmoq, shtift, shpilka, shplint* va h.k., shuningdek, ushbu leksema boshqa so'z turkumi sifatida ham turli ma'nolar kasb etadi: *hayotini buzmoq, so'zidan ilib olmoq* va h.k.

Olimlar fan va texnikaga oid terminlarni standartlashtirishga oid fikrlarni beradilar. Muayyan terminning iste'molga kirib kelishida uning yaratuvchisi yoki ijtimoiy hodisalari ta'sirida ba'zi terminlarning o'rniga boshqa birining qo'llanish holatlarini inkor etib bo'lmaydi. Lingvistik ta'minotda terminlar lug'atining nisbatan to'liq berilishi tarjima sifatini oshirishga xizmat qiladi.

Ilmiy matnlarning leksik birliklari quyidagi razryadlarga ko'ra bo'lishimumkin:

- Adabiy tilda qabul qilingan so'zlar: *predlog, yordamchi so'zlar, olmosh*
- Umumadabiy tilda qabul qilingan, qoidaga ko'ra ilmiy matnda muayyan ma'noga ega so'zlar: *qism, shart, qila olmoq*
- Ilmiy bo'lmagan matnlarda juda kam uchraydigan va aynan ilmiy matnga xos xarakterni ko'rsatuvchi so'zlar: *tahlil qilmoq, tasniflamoq, metod, neyro*
- Frazologik birliklar: ehtiyojni qondirmoq, komponentlarningta'sirlashuvi...
- Muayyan fan doirasida maxsus terminologik birliklar: *ehtimollik, to'plam, formal til*
- Ilmiy matnda uchramaydigan, biror muayyan fanning predmeti sifatida umumadabiy tilda qabul qilingan birliklar: *aborigen, absurd*

ADABIYOTLAR TAHLILI VA METODOLOGIYA

Mashina tarjimasini yaratishda mashina tarjimasini zamonaviy tilning qancha leksik boyligini o'z ichiga qamrab olishi ko'pchilikni qiziqtiradi. Xalqaro

tashkilot INFOTERM (Avstriyaning Vena shahridagi Xalqaro terminologik markazi) bergan ma'lumotlarga tayaniladigan bo'lsa, Yevropa tillariga oid terminlar soni 50 million, mahsulot (tovar) nomlarining miqdori 100 millionga etadi. Bu shuni anglatadiki, jamiyatda yaratiladigan ishlab chiqarilgan har qanday mahsulot yokibiror yangilik muayyan nomga ega bo'ladi va termin sifatida leksikografiyani boyitadi.

Logos (LOGOS), Engspan (ENGSPAN) hamda Sistran (SYSTRAN) mashina tarjimasini tizimlari transfer usulidan foydalanadilar. 1964-yili Koliforniyada tashkil etilgan Sistran kompaniyasining tarjima tizimida 27 til kombinatsiyasi mavjud. So'zlar umumiy hamda texnikaviy ma'noga (termin sifatida) ega bo'lib, so'z kombinatsiyalarining sintaktik va semantik tahlili asosida tashkil topgan. Semantik tasniflangan tizim o'zining ierarxik strukturasi 500 ta kategoriyani qamrab oladi. Eng murakkab vazifa bu so'zning ko'p ma'noliligidir. Sistranda tarjima jarayonining transfer bosqichida maxsus dasturlar ishlab chiqilgan. Unga ko'ra dastlab tahlilda so'zning jonli yoki jonsiz ekanligi aniqlangan. Lug'atda biror so'z [+jonli / jonsiz] tarzida kodlangan: (masalan, *seal* (1) *tyulen*; 2) *muhr*, *tamg'a*), *acquaintance* (1) *tanish*; 2) *tanishish* kabi). Mavzuga oid glossariylarda tarjima qilinayotgan so'zning ikkinchi tilda bir nechta ma'nolarda ifodalanishi mumkin. Jumladan, ingliz tilida *stem* termini bir nechta sohalarda qo'llanilib, ikkinchi tilda turlicha tarjima qilinishi mumkin: botanikada *plant stem* - "o'simlik ildizi", anatomiyada *brain stem* - "miya tomiri", tilshunoslikda *word stem* - "so'z negizi" kabi. Bu tizimda foydalanuvchi tarjimadan avval o'nta sohadan birini tanlab, so'ng undan foydalanadi. Shunda har birsoha uchun so'zlarga kiritilgan kod ham tanlanadi. Bunday masalalarni hal etishda otli so'z birikmalar shaklida tuzilgan katta korpuslar ahamiyatga ega. Bundan tashqari LOGOS mashina tarjimasini tizimi salmoqli ishlarni amalga oshirgan. Bu tizimda semantik noaniqliklarni hal etishga mo'ljallangan turli sohalarga doir besh yuzdan ziyod maxsus ichki lug'atlar kiritilgan. Unda 130 dan ortiq ierarxik tarzda tuzilgan semantik kategoriyalar mavjud. Shuningdek, VISTA (inglizcha-ruscha, ruscha-inglizcha) mashina tarjimasini tizimi 3 mln. 300 ming leksik birliklarni o'z ichiga oladi. Bularning 80 % ini uzunligi ikkitadan o'n yettitagacha bo'lgan so'z birikmalari tashkil etadi. Bunga qo'shimcha qilingan tematik lug'atlarning hajmi 400 mingdan ortiq leksik birliklardan tashkil topgan [2]. Mashina tarjimasini uchun terminologiya lug'atini lingvistik ta'minotga kiritish, ularning umumiy va xususiy ma'nolarini kodlash orqali ingliz tilidagi matnlarni o'zbek tiliga tarjima qilish imkoniyati kengayadi. Buning uchun tanlab olingan sohalar doirasida ikki tilning terminologik lug'atlarini dasturiy ta'minotga kiritish, ularning semantik tahlilini berish hamda terminlarni standartlashtirish lozim bo'ladi. Chunonchi, ijtimoiy- gumanitar fanlar bo'yicha terminlarning bir xilligini ta'minlash maqsadga muvofiq. Bu masala yuzasidan Xalqaro

standartlantirish tashkiloti (ISO-International Standartization Organization) tomonidan terminlarni standartlashtirish masalalari o'rganilib kelinmoqda. Shu o'rinda bu kabi ajratishning nazariy asoslari yana bir qatorasarlarda o'z tasdig'ini topganligini e'tirof etamiz [3].

A.A Reformaskiyning nuqtai nazaricha, terminologiya og'zaki ifodada mos ravishda mustahkamlangan, muayyan fanning tizimlashtirilgan tushunchasidir [4]. Mashina tarjimasida terminologiyani alohida soha sifatida o'rganish shu jihatdan zarur hisoblanadi. Terminlarning yasalishi uch usul bilan amalga oshiriladi: semantik, morfologik, sintaktik terminlarning semantik usul bilan yasalishiga metoforik va metonimik ma'nolarning boshqa terminologiya bilan bog'liq holda xususiylashishi xos. Bunga transpozitsiya va transmutasiya hodisalarini keltirish mumkin. Morfologik usul bilan termin yasashda suffiksasiya, prefiksasiya, konversiya va asosni o'zlashtirish hodisalari yetakchilik qiladi. Sintaktik usul bilan termin yasashda transformatsion usulning o'rni sezilarli. Bundan tashqari konkretizatsiya, generilizatsiya, kompensatsiya, transliteratsiya, eksplikatsiya, amplifikatsiya, ma'no differensiyalashuvi va h.k. hodisalarni ham nazardan qochirmaslik kerak [5]. Birgina mashhur inglizcha Vebster lug'atidagi 600 000 so'zdan 500 000 tasining mutaxassislikka oid leksemalardan tashkil topgani terminologiyani chuqur tahlil etishni taqozo etadi [6]. Zero, ilmiy-texnikaviy terminlar jamiyatning barcha sohalarida qo'llaniladi. "London Computer Integrated Translation GmbH" (1987) ko'rsatgan ma'lumotlarga ko'ra, dunyo bo'yicha ilmiy-texnikaviy matnlarni tarjima qilish hajmi yiliga 200 mln. sahifani tashkil etadi, bu esa har yili 15% oshib borayotganligini tasdiqlaydi [7]. Shu o'rinda ilmiy matnlarni kompyuter yordamida tarjima qilish mazkur jarayonni optimallashtirishga ko'mak beradi. Ilmiy-texnikaviy matnlarni tarjima qilishda uning oldida turgan mas'uliyatli vazifalar borligini ko'rsatadi.

NATIJARLAR

O'zbek tilida mashina tarjimasidagi asosiy masalalardan biri bu terminlarning tarjimasini to'g'ri amalga oshirishdir. Lingvistik ta'minotda terminlar lug'atining to'liq berilishi tarjima sifatini oshirishga xizmat qiladi. Bundan tashqari soha mutaxassislari uchun ham kompyuter texnologiyalardan foydalanishning bir qator afzalliklari bor. Chunki termin sifatida yangi kirib kelgan so'zlarning tarjimasini to'g'ri tushinishda hamda muayyan terminning bir necha sohada qo'llanib, turlima'no kasb etgani bois ularning lingvistik bazasini yaratish uning nechog'lik dolzarbligini ko'rsatadi.

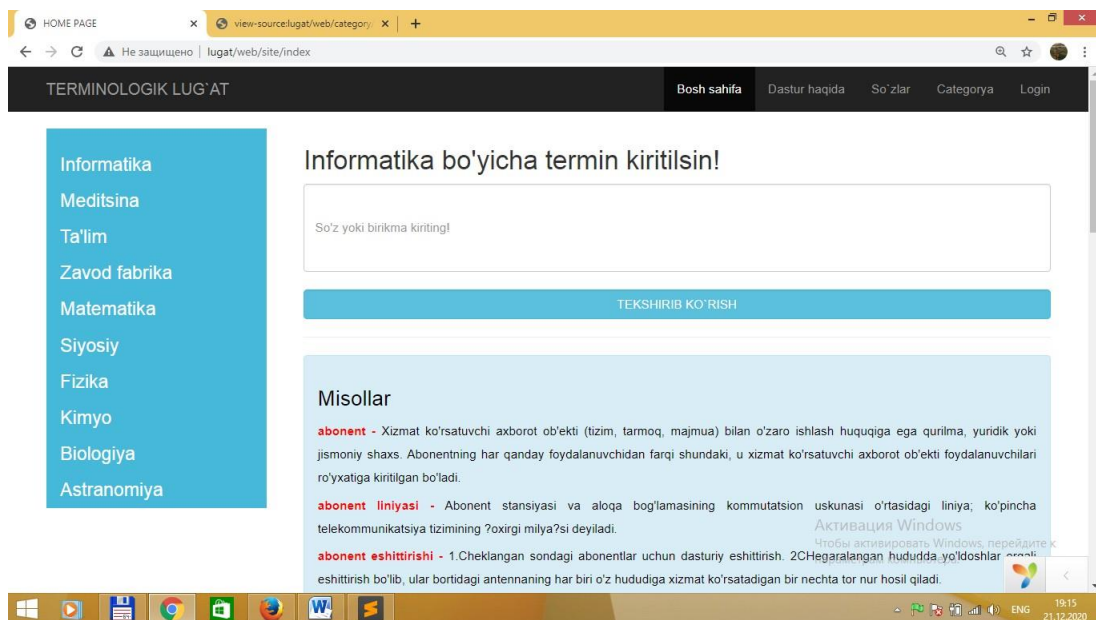
Ochiq manbali terminologik ma'lumotlar bazasini loyhalashtirishdan maqsad fan va sohalarni o'rganishda terminologik ma'lumotlarning umumiy bazasi yo'qligi bir qator qiyinchiliklarga olib keladi. Undan tashqari mashina tarjimasida terminlarning tarjimasini to'g'ri amalga oshirish, terminlar lug'atining to'liq berilishi tarjima sifatini

oshirishda katta ahamiyatga ega. Shu bois biz qilayotgan loyihada oldimizga qo‘ygan asosiy maqsadlarimiz fan va sohalarni umumiy terminologik ma’lumotlar bazasini loyhalashtirish, barcha uchun ochiq manbali bo‘lishini taminlash, terminologik ma’lumotlar bazasidagi ma’lumotlarni doimiy ravishda yangilab borish, o‘zgartirish va himoyalash imkoniyati mavjud bo‘ladi.

Ochiq manbali terminologik ma’lumotlar bazasini loyhalashtirishda quyidagi ishlarni bajarishimiz kerak bo‘ladi:

1. Fan va sohalarga oid terminologik adabiyotlardan foydalangan holda boshlang‘ich terminologik ma’lumotlar bazasini yaratib olishimiz lozim;
2. Internetda onlayn qidiruv tizimidan foydalanib ma’lumotlar bazasini boyitib borishni yo‘lga qo‘yish;
3. Fan va sohalar mutaxassislari o‘z sohasining yangi terminlarini bazaga kiritish imkoniyatini yaratilishi;
4. Terminologik ma’lumotlar bazasini ochiq manbali qilish imkoniyatlarini ko‘rib chiqish;
5. Ochiq manbali terminologik ma’lumotlar bazasini aniqliligini oshirish va uni himoyalash.

Quyida yaratilgan loyihaning umumiy ko‘rinishi keltirilgan.



Platforma quyidagi bo‘limlardan iborat:

- ✓ Bosh sahifa – Bosh sahifada biron - bir fan yoki soha tanlanadi va shu bo‘yicha terminlarni izlash imkoniyati mavjud bo‘ladi.
- ✓ Dastur haqida – dastur haqida umumiy ma’lumotlar keltiriladi.
- ✓ So‘zlar – bu bo‘limda terminlarning umumiy bazasi joylanadi. Shuningdek,

yangi terminlarni foydalanuvchi tomonidan kiritish vazifasi yuklanadi.

✓ Kategoriya – yangi fan yoki sohalarni qo‘shish uchun mo‘ljallangan.

✓ Login – bu bo‘limda berilgan login porollari bilan soha mutaxassislari terminlar bazasini kuzatib borish va yangi terminlarni qo‘shish imkoniyati beriladi.

Fan va sohalarga oid terminologik adabiyotlardan foydalanib ma‘lumotlar bazasini yaratishda adabiyotlarda berilgan terminlarni aniqliligini tekshirish va aynan shu sohaga mansub bo‘lgan adabiyotlardan foydalaniladi.

MUHOKAMA

Fan va sohalar mutaxassislari o‘z sohasining yangi terminlarini bazaga kiritish uni kuzatib borish imkoniyatini yaratishdan oldin uni aynan o‘z sohasining mutaxassisi ekanligini tekshirib olish zarur. Buning uchun uni shu sohaga tegishli bo‘lgan savol–javob (test va yozma ravishda) yoki avtomatik suhbat asosida bazani kuzatish va ma‘lumot kiritishga ruxsat beriladi. Soha mutaxassisi bazaga ma‘lumot kiritganda imkoni boricha qiyinchilikka duch kelmasligiga e‘tibor beriladi va ma‘lumot kiritishda osonlashtirilgan platformasi ko‘zda tutiladi.

XULOSA

Ochiq manbali terminologik ma‘lumotlar bazasini aniqliligini oshirish maqsadida izlanyotgan terminda keltirilgan ma‘lumotlarini aniqligi haqida turli xil qiziqarli so‘rovnomalarni o‘tkazish va so‘rovnomalar natijasidan kelib chiqib shu termin haqidagi ma‘lumotlar bazasini umumlashtirish va uni himoyalash imkoniyati tug‘iladi.

Umid qilamizki, ushbu Ochiq manbali terminologik ma‘lumotlar bazasi fan va sohalar bir qator yengilliklar yaratishda, fanlarni mukammal o‘rganishda, ya‘ni terminlar ma‘nosini va mohiyatini tushinishga imkoniyat yaratadi. Shuningdek, ushbu platforma barcha fan va sohalar bo‘yicha terminologik ma‘lumotlar bazasi sifatida foydalanishga imkon yaratadi.

REFERENCES

1. Марчук Ю. Компьютерная лингвистика (2007). –С. 196.
2. <http://www/retrans.ru/publikacii/statyi/MT>
3. М.Н. Литвинова (2012) Практикум по переводу (английский язык): учеб.-метод.пособие / авт.-сост. М.Н. –С. 74.
4. Реформацкий А.А. (1961) Что такой термин и терминология? / Вопросы терминологии. –С. 49.
5. Онанченко А.В. (2001) Способы образования переводческих терминов в русском языке. –С. 25
6. Webster’s Third New International Dictionary (1961) Webster 3.

7. Гул Хасан (2007) Лексико –семантические особенности перевода научно – технического текста с языка урду на русский язык: Дисс. ... док-ра филос.наук. – Исломобод,–С.18.
8. Abdurakhmonova N. (2020) Kompyuter lingvistikasi (darslik). Globeedit, 395 b.
9. Абдурахмонова Н. (2020) Замонавий лингвистик корпусларнинг компьютер моделлари Ўзбекистонда хорижий тиллар (1/30), 50-58
10. Abdurakhmonova N.(2019) A two-level morphological analysis of the Uzbek corpus Материалы IV международного научного конгресса иностранная филология. Социальная и национальная вариативность языка и литературы), 425-431
11. Abdurakhmonova N. (2016) The bases of automatic morphological analysis for machine translation Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta (2/38), 12-17.
12. Хакимов М., Абдурахмонова Н. (2013) Логико- лингвистические модели слов и предложений английского языка для многоязычных ситуаций компьютерного перевода Компьютерная обработка тюркских языков. Труды, 292-302