

SUN’IY INTELLEKT ALGORITMLARI ASOSIDA O‘ZBEK TILIDA MASHINALI
TARJIMANI AMALGA OSHIRISH MUAMMOLARI

Omonturdiyeva Mahfuza Karim qizi

Muhammad al-Xorazmiy nomidagi Toshkent Axborot Texnologiyalari universiteti “Sun’iy intellekt” kafedrasida magistranti, Muhammad al-Xorazmiy nomidagi Toshkent Axborot Texnologiyalari universiteti, Toshkent, O‘zbekiston.

mahfuzaomonturdiyevadev@gmail.com

<https://doi.org/10.5281/zenodo.7856092>

Abstract. *In this article, the stages of implementation of machine translation based on artificial intelligence algorithms, the current results in this regard are analyzed, and the relevance of the creation of Uzbek language translator programs is highlighted.*

Keywords. *artificial intelligence, machine translation, words, phrases, language grammar, translation, Google Translate, Microsoft Translator, DeepL, Yandex Translate, Amazon Translate, parallel corpora.*

Kirish. Sun’iy intellekt algoritmlari yordamida tarjima dasturi yaratish, turli til ma’lumotlar bazasini va algoritmlarni birgalikda ishlatish, test jarayonidan o‘tkazish va dasturni optimallashtirishni o‘z ichiga olgan bir jarayon bo‘lib, sodda va oson ishlatiladigan bir interfeys orqali bir tilni boshqa tilga o‘girish imkoniyatini beradi.[1]

Mashinali tarjima, bugungi kunda ko‘p tilni o‘rganish va tarjima qilish uchun odatiy usuldir. Mashinali tarjimada har qanday so‘zning ma’nosi, konteksti va sinonimlari, barcha tillarning grammatik xususiyatlari ham o‘rganiladi. Shuningdek, bir nechta tillarda tuzilish jihatdan kamchiliklar ham kuzatilishi mumkin.

Bunday muammo va ziddiyatlar bilan bog‘liq yechimlardan biri - Sun’iy intellekt algoritmlari asosida mashinali tarjimani amalga oshiruvchi dasturiy vositalar yaratishdir. Bu sun’iy intellektning bir qismini tashkil etadi va ko‘p tilni o‘rganish va tarjima qilishni osonlashtirish uchun yaxshi imkoniyatlar beradi.

Bu xizmatning ishlashi esa oldin matnni tahlil qilishdan iborat bo‘ladi, keyin esa ushbu matnni ma’nosi bo‘yicha qayta yozib berishdan. Bu narsa sun’iy intellektning bir nechta shakllarini qo‘llash, ma’lumotlar tuzilishi, tahlil qilish va tarjima qilish uchun xizmat qilish kabi yaxshi imkoniyatlar beradi.

Biroq, sun’iy intellekt asosida mashinali tarjima kamchiliklarga ega. Hozirgi tarjima dasturiy vositalarida dastlabki xatolar, tuzilish bo‘yicha ko‘rsatkichlar va aniqlik darajasini saqlashda bo‘lgan muammolar hamda tilning grammatik va ma’no jihatdan ko‘rsatkichlaridan tashkil topadi. Shu sababli, Sun’iy intellekt asosida mashinali tarjima dasturiy vositalarini yaratishning o‘ziga xos muammolari bor, lekin katta imkoniyatlar ham mavjud. Yaqin yillar ichida, bu sohada yana ko‘p ishlar kuzatilmoqda va mashinali tarjimani oson va aniq qilish uchun yaxshi dasturiy vositalar ishlab chiqarilmoqda.

Mashinali tarjimani amalga oshirish bosqichlari quyidagilardan iborat bo‘ladi:

1. Ma’lumotlarni to‘plash: Tarjima uchun kerakli til to‘plangan va o‘zaro bog‘liq ma’lumotlar (so‘zlar, jumlar, iboralar, matnlar, matn segmentlari, tarjimalar) topiladi va ma’lumotlar tahlil qilinadi.

2. Tahlil: Sun'iy intellekt algoritmlari, tahlil qilish uchun matnlarni va tilni tahlil qilish uchun turli usullar ishlatadi. Misol uchun, sintaksis va morfologik tahlil, kalit so'zlar, to'plam, so'zlar, iboralar, bog'lovchi so'zlarni va boshqa xususiyatlarni tahlil qilinadi.

3. Tarjima qilish: Sun'iy intellekt algoritmlari, tahlil natijalariga asosan tarjima qilishni amalga oshiradi. Tarjima tizimlari, ma'lumotlar, til qoidalari va iboralar kabi faktorlar bilan ham ishlaydi.

4. Qiymatlash: Tarjima natijalari, qiymatlash jarayonidan o'tkaziladi va nihoyat, qulay va sodda tarjimani taqdim etish uchun to'g'ridan-to'g'ri formatga o'tkaziladi.

5. Taxmin qilish: Mashinali tarjima tizimlari tarjima natijalarini taxmin qilishga yordam beradi. Misol uchun, ma'lumotlarni kuzatib borish, mavjud ma'lumotlar asosida turli til qoidalari, masalan, til uchun mo'ljallangan algoritm, neyron tarmoqlar va boshqa qoidalarga asoslangan usullar yordamida natijalarni taxmin qilish mumkin.[2,4]

Quyida sun'iy intellekt algoritmlari asosida yaratilgan mashinali tarjima dasturlari bilan tanishishingiz mumkin.

- Google Translate: Googlening onlayn tarjima dasturi, dunyoning eng mashhur tarjima tizimlari qatoridan joy olgani bilan mashhur. Bu tizim, ko'plab tillarga, shu jumladan inglizchaga, ispanchaga, fransuzchaga, nemischaga va yaponchaga tarjima qilish imkoniyatiga ega.

- Microsoft Translator: Microsoftning onlayn tarjima tizimi, eng yaxshi mashinali tarjima dasturlaridan biri sifatida tanilgan Bu tizim, turli xil tillarni tarjima qilish imkoniyatiga ega.

- DeepL: DeepL yangi bir tarjima dasturi hisoblanadi. Bu tizim, ko'plab tillarni qo'llab-quvvatlaydi va "neyron tarjima tarmoqlari" deb nomlanadigan usulni ishlatadi.

- Yandex Translate: Yandexning onlayn tarjima dasturi, rus, ingliz va turk tillariga, shuningdek, ko'plab boshqa tillarga tarjima qilish imkoniyatiga ega.

- Amazon Translate: Amazonning onlayn tarjima tizimi, ko'plab xususiyatlarga ega, shu jumladan kompaniyaning "Amazon Web Services" (AWS) jihatidan qo'llab-quvvatlanadi.

- Apertium: Apertium, ochiq manba kodli onlayn tarjima tizimi hisoblanadi va turli xil tillarni tarjima qilish imkoniyatiga ega.

O'zbek tili tarjima qilish bo'yicha bir qancha qiyinchiliklarga ega:

- O'zbek tili grammatik jihatdan aniq qoidalarga asoslangan til va unda so'zlovchi tomonidan gapirilgan gapda yashirin grammatik qo'shimchalar tushib qolishi ko'p kuzatiladi;

- O'zbek tilida sinonim va antonimlar juda ko'p bo'lib ularning gapda qo'llanilish jihatidan ajratib olishga to'g'ri keladi;

- Iboralarni tarjima qilish dasturiy jihardan qiyinchilik tug'diradi.

Yuqoridagi muammolar o'zbek tiliga tarjima qilish uchun dasturlarga qiyinchilik tug'diradi. Bunday kamchiliklar yuqorida sanab o'tilgan dasturlarda ham kuzatilishi mumkin.

O'zbek tili tarjimon dasturlarining yaratilish dolzarbligi, dasturlar tuzish usullari va yaratuvchilarning ma'lumot va tajribalari bo'yicha o'zgarishi mumkin. Sun'iy intellekt asosida tuzilgan dasturlar, o'zbek tiliga tarjima qilish uchun turli xil qoidalarga ega bo'lishi kerakligini o'rganishdan boshlanadi.

**INTERNATIONAL SCIENTIFIC AND TECHNICAL CONFERENCE
“DIGITAL TECHNOLOGIES: PROBLEMS AND SOLUTIONS OF PRACTICAL
IMPLEMENTATION IN THE SPHERES”**

APRIL 27-28, 2023

O‘zbek tili turk tillari oilasiga kiradi va bu tillarning o‘ziga xos grammatikasi va so‘zlar mavjudligi, shuningdek, o‘ziga xos mazmun va ma’nolar ko‘rsatishi sababli, o‘zbek tili tarjimon dasturlarining yaratilishi juda qiyin va kerakli vaqt va mablag‘lar katta bo‘lishi mumkin.

Dasturlarni tuzishda, tarjima uchun kerakli o‘zbek tili so‘zlarining va gaplarining ma’nolari, ko‘rsatilgan ma’nosini to‘g‘ri tarjima qilish mumkin bo‘lgan boshqa tildagi so‘zlar va gaplar bilan taqqoslanishi kerakligi, bu esa ma’lumotlar bazasining to‘g‘ri tayyorlanishi va dastur ishlashining samaradorligi uchun zarur bo‘lishi mumkin.

Bunday ma’lumotlarni to‘plash bugungi kunda parallel korpuslar muammosiga kiradi.

Parallel korpus - bu ikki yoki undan ortiq tildagi matnlar to‘plami bo‘lib, ular jumla yoki gap darajasida mos keladi, shuning uchun tegishli jumlar yoki gaplarni osongina solishtirish mumkin. Parallel korpus ko‘plab tabiiy tillarni qayta ishlash vazifalarida, jumladan, mashina tarjimasi, tillararo ma’lumot qidirish va ikki tilli leksikografiyada qo‘llaniladi. Turli tillardagi matnlar orasidagi yozishmalarni tahlil qilib, mashinali o‘qitish modellari ular o‘rtasida to‘g‘ri tarjima qilishni o‘rganishi mumkin. Parallel korpus qo‘lda yoki avtomatik tarzda yaratilishi mumkin va ko‘p tilli tabiiy tillarni qayta ishlash bo‘yicha ishlaydigan tadqiqotchilar va ishlab chiquvchilar uchun muhim manba hisoblanadi.[5]

“Toshkent lingvistik korpus”ida o‘zbek va rus tillarida 2000 ga yaqin parallel gaplar mavjud.

O‘zbek-ingliz parallel korpusida 50 000 ga yaqin parallel gaplar mavjud.

Bu raqamlar yangi parallel korpuslar paydo bo‘lishi yoki mavjud korpuslar kengaytirilishi bilan o‘zgarishi mumkin.

O‘zbek tilidagi parallel korpuslar, hozircha yetarlicha katta emas, ammo bu borada izlanishlar olib borilmoqda.

"National Corpus of Uzbek" (O‘zbek milliy korpusi): O‘zbek tili bo‘yicha eng katta korpus hisoblanadi va matnlarining bir qismi rus va ingliz tillarida ham mavjud. Bu korpusda tilning ko‘plab turlari va xususiyatlari keng ko‘rib chiqilishi mumkin.

Ko‘rinib turibdiki, parallel korpuslarni yaratish bizdan ko‘p vaqt va mablag‘ talab qiladigan jarayon. Bunday tarjimon dasturlarini tuzish uchun esa bizga katta hajmdagi parallel korpuslarni ishlash kabi qiyinchiliklarga duch kelishimizni bildiradi. Parallel korpuslarni yaratish uchun tarjima qilish uchun kerakli xususiyatlar, o‘zbek tilining grammatikasi va tarixi haqida tushunchaga ega bo‘lish zarurdir. Bunday tushunchaga ega bo‘lgan yaratuvchilar, o‘zlarining dasturlarini yaratishda ma’lumotlarni qo‘shish va tarjima natijalarini tekshirish uchun ilg‘or dasturlar ham ishlatishadi.

Xulosa: Sun‘iy intellekt algoritmlari asosida mashinali tarjima amalga oshirish uchun bir nechta bosqichlar kerak bo‘ladi, shu jumladan, ma’lumotlarni to‘plash, tahlil qilish, tarjima qilish, qiymatlash va taxmin qilish. Hozirgi kunda ishlatilayotgan mashinali tarjima dasturlari sun‘iy intellekt algoritmlari yordamida yaratilgan va turli xil tillarni tarjima qilish imkoniyatiga ega. Lekin ular o‘zbek tilini tarjima qilish bo‘yicha kamchiliklarga ega. Bu kamchiliklarni bartaraf etish uchun bunday dasturlar shu tilni grammatik jihatdan yaxshi bilgan dasturchilar tomonidan yaratilishga zarurat tug‘iladi. Shuning uchun mazkur tezis mavzusi aynan o‘zbek tilini tarjima qiluvchi mashinali tarjimonni amalga oshirish muammolarini tahlil qilishga bag‘ishlandi.

REFERENCES

**INTERNATIONAL SCIENTIFIC AND TECHNICAL CONFERENCE
“DIGITAL TECHNOLOGIES: PROBLEMS AND SOLUTIONS OF PRACTICAL
IMPLEMENTATION IN THE SPHERES”**

APRIL 27-28, 2023

1. Sebastian Raschka, Vahid Mirjalili Python Machine Learning. Published by Packt Publishing Ltd. December 2019. p.1-13
2. Salvador Ordorica, Why Technology Will Not Replace Professional Translators.Forbes. Oct 26 2020
3. Manuel Herranz, What Are the Main Advantages and Disadvantages of Machine Translation. Pangeanic
4. Kajetan Malinowski with Janette Mandell, The Future of Language Technology: The Future of Machine Translation. Lionbridge January 22, 2021
5. Sylviane Granger, Comparable and translation corpora in cross-linguistic research, Centre for English Corpus Linguistics Université catholique de Louvain. Researchgate 2010