Global Scientific JOURNALS

GSJ: Volume 8, Issue 12, December 2020, Online: ISSN 2320-9186
www.globalscientificjournal.com

ALGORITHM BASED ON LINGUISTIC MODELS IN MACHINE TRANSLATION BETWEEN ENGLISH AND UZBEK

Xolisa Axmedova
Tashkent State University of the Uzbek Language and literature named after Alisher Navoi
xolisa9029@mail.ru
Dilfuza Yusulova
Tashkent State University of the Uzbek Language and literature named after Alisher Navoi
dilfuzayusupova58@gmail.com
Manzura Abjalova
Tashkent State University of the Uzbek Language and literature named after Alisher Navoi
Manzura_ok@mail.ru

ABSTRACT

The article is devoted to the analysis of simple sentences' structure of English and Uzbek languages. We propose an algorithm that solves crucial problem for machine translation of these unrelated languages, and the linguistic database that gives the possibility to implement the process of machine translation.
Keywords: database, machine translation, tokenization, programming and linguistic database, algorithm.

Computational linguistics is one of the complicated fields which crossroads of linguistics and computational technologies. Because it links directly with natural language processing, indeed it also depends on several factors that are psychological, cognitive, and cultural and so on. Nevertheless, translation is not only technical process but also creative activity that based on including both material and mental capability of human being. Therefore, for machine translation it is important to identify what kind of texts would be objects in the automatic process. We clarify the text in terms of genres like official or scientific texts that are more formal than others are. However, a lot of breakthrough in the field involving oral and written form of all genres texts since many attempts have been implemented over the world.

A new technology in the field of machine Translation in Computer Modeling (TCM) has been added. This technology provides a high level of machine translation, that is, between 65 and 80% of translations between the two languages. It includes a 7-step process. At the first stage - syntactic and semantic analysis of natural languages in the TCM -adapted method and semantic bases of natural languages were created; At the second stage- logical and linguistic models of words, words and phrases in the natural language were created; at the third stage - an expanded introductory language for mathematical modeling of natural languages; in the fourth stage - mathematical models of speech, word brochures and extensible introductions based on logical linguistic models of natural language; at the fifth step- it is required to create databases on the common and natural language vocabulary dictionaries, and in the subject areas two or more natural language dictionaries; at the sixth stage - the creation of algorithms related to the requirements of

multilingualism based on mathematical models of natural language; the seventh stage is the creation of a software and interpreter environment based on the results obtained.

Regarding progress, today as we mention some approaches of machine translation like neuro machine translation, statistical, phrasal-based etc. Owing to globalization and interactive communication between nations in Internet, translation tools have a pivotal role to ease and make the atmosphere that is necessary and so fast with quality to take daily information and transform them consumer as soon as possible. It is not even in social networking, but exchange academic background at any time at different parts of the world gives a great chance to analyze and criticize them wherever its needed. Therefore, in machine translation the Uzbek language is important as it one of Turkic language.

Our article is focus how to build up algorithm for machine translation from English into Uzbek and vice versa.

Firstly, it is applied morphological analysis in the first stage: tokenization (take apart word form) -> lemmatization (the analysis of morphemes) -> stemming (identify the roots of the words). Thereafter syntactic models of the text compared and checked each other.

Obviously, database is well structured systematically and by structure to keep data that are used in urgent time accurately and properly which are asked somehow. It is should be input symbols for environment of machine translation.

Table 1.

Data Name	Function
R_i	The database of phrase and terms of the scientific spheres.
Q1	The database of all of the words root in the language.
K1	The database of all derivational words
V2	Clause elements
V3	The database of parts of speech

The environment translation services for scientific text. It is very important to address Grammar of the languages so that to identify the structure of the sentence and parts of speeches in the text. It could do this work through two directions: English-Uzbek, Uzbek-English.

Firstly, dividing into several parts of speech of input text (Z) and each words are taken the other term database; they are replaced in terms of grammar. We display the functional chart of translation algorithm:

The following symbols input in the entry part of language in order to model of natural language:

T3i1-translation into other language and the massive including the function in the sentence,$1 \le i1 \le m$;

T4j1- translation into other language,$1 \le j1 \le m1$;

T2-translated text; E4-subject; G2 -predicate; E5- attribute;

E6-object; E7- modifier.

There are two appropriate models of sentence in both of languages.

a) the different mathematical models of types of indicative mood in Uzbek:

I. 1. <E4>↓⊕<E5>↓⊕<E6>↓⊕<E7> ⊕<G2>.

2. ↓<E5> ⊕<E4>↓⊕<E6>↓⊕<E7> ⊕<G2>.

3. ↓<E5>↓⊕<E5> ⊕<E4>↓⊕<E6>↓⊕<E7> ⊕<G2>.

4. <E4>↓⊕<E5>↓⊕<E6> ⊕<G2>.

5. <E4> ⊕<G2>.

6. <E4>↓⊕<E7> ⊕<G2>.

7. <E4>↓⊕<E6> ⊕<G2>.

Thus we apply a bit change of mathematical models which presented at [1,3,4] types of

component of sentence. Hence, some exact parts of speech could be appropriate clause elements in some cases that identified as models of the text. Afterwards it is taken from other translation in the second language and it is replaced in order by normal principles. In next stage algorithm takes function in order to the most optimal and meaningful translation. Above mentioned the forms Uzbek sentences are formed as English mathematical models:

1. <E4>↓⊕<E5> ⊕<G2>↓⊕<E6>↓⊕<E7>.
2. ↓<E5>⊕<E4>↓⊕<E7> ⊕<G2>↓⊕<E6>.
3. ↓<E5>↓⊕<E5>⊕<E4> ⊕<G2>↓⊕<E6>.
4. <E4> ⊕<G2>↓⊕<E6>↓⊕<E5>.
5. <E4> ⊕<G2>.
6. <E4> ⊕<G2>↓⊕<E7>.
7. <E4> ⊕<G2>↓⊕<E6>.

b) Let's take the mathematic models of simple interrogative sentences of Uzbek language as an example:

1.<M4>↓⊕<E5>↓⊕<E5>↓⊕<E6>⊕<G2>
2. <M4>↓⊕<E6>↓⊕<E5>⊕<G2>
3. ↓<E6>⊕<M4>⊕<G2>
4. <M4>↓⊕<E5>↓⊕<E6>⊕<G2>

These interrogative sentences suit in English such models as following examples:

1. < M4 >⊕<G2>↓⊕<E7>↓⊕<E5>↓⊕<E6>
2. < M4 >⊕<G2>↓⊕<E6>
3. < M4 >⊕<G2>↓⊕<E5>↓⊕<E6>
4. < M4 >⊕<G2>↓⊕<E6↓⊕<E7>

Using above mentioned database structure of sentences and terms, translation algorithm is given like this:

Q1_uz=» SELECT * FROM `Q1_uz`»-all stems in Uzbek;
K1_uz=» SELECT * FROM `K1_uz`»-all word forms in Uzbek;
Q1_ru=»SELECT * FROM `Q1_ru`»- all stems in English;
K1_ru=»SELECT * FROM `K1_ru`;»- all word forms in English;
Ei – sentence taken from text Z, 1≤i≤n; L1j – words taken from Ei , 1≤j≤n1;

After doing algorithm [2], the following "search" algorithm divides into Z sentences, and after that it breaks apart words or word combinations, then each word formations is searched in the database of stem list, if there is not need words turning another one type of database. After finding words, taken translation form the target language. As we take one more example for Uzbek-English direction the 1st translation algorithm like this:

1. Search the words in L1j from Q1_uz. If find go 2nd step, otherwise 4th step;
2. Take the stem from Q1_uz in terms of English order (ID);
3. Take translation of stream of Q1_ru and go through the 7th step;
4. Search each word in L1j from K1_uz;
5. Take the order (ID) word formation in K1_ru form K1_uz;
6. Take translation of word formation from K1_ru;
7. Identify the function in the sentence and replace in the massive T3i1;
8. Pass filled massive of T3i1 to function UzbekEnglish (T3i1);
9. Replace the results of function UzbekEnglish (T3i1) to T2;

Here UzbekEnglish(T3i1) [2] function which is written translation algorithm for Uzbek-English direction. UzbekEnglish(T3i1) function is written as following. So we used some signs to write

function:

1. ET3k1 –Uzbek and English the structures that are suited each other 1≤ k1≤m2;
2. Load the functions of words which are input T3i1 to E8k massive;
3. Find appropriateness structure sentence to E8k form ET3k1;
4. Take found the fords as clause elements from ET3k1 and load to T2;

This function is such a form in programming language (in Java):

```
private String  UzbRus(String          suz)     throws
ObjectNotFoundException { int rusId=0; String rusSuz =””; int gapBulagiId=0;
U z a k S u z l a r u s = u z a k S u z U z b e k D a o . getUzakUzbekByWord(suz);
if(us.getUzakSuzlar().equals(suz)){ rusId=us.getUzakEnglishId();
List<UzakEnglish>ueList=uzakSuzEnglishDao. getuzakSuzlarListByRId(rusId);
for (UzakEnglish ue : ueList) { rusSuz=ue.getUzakEnglish();
}}else{YasamaSuzlar ys=yasamaSuzUzbekDao. getYasamaUzbekBySuz(suz);
if(suz.equals(ys.getYasamaSuzlar())){
rusId=ys.getYasamaEnglishId();
YasamaEnglish ye=(YasamaEnglish)
yasamaSuzEnglishDao.getYasamaEnglishListByRId(rusId);
rusSuz=ye.getYasamaEnglish();
}else{ rusSuz=suz;}}return rusSuz;}
```

The algorithm 2 is for English-Uzbek direction like this:

1. Search each word in L1j from Q1_rus. If it is found, go to the 2nd step, otherwise to the 4th ;
2. Take the order (ID)stem in English from Q1_rus;
3. Take translation stem from Q1_uz and go to the 7th step;
4. Search each word in L1j from K1_rus;
5. Take the order (ID) in word formation in K1_uz from K1_rus;
6. Take translation derivative word from K1_uz;
7. Identify the function of the word in the sentence and replace in the massive of T3i1;
8. Pass filled massive T3i1 to function EnglishUzbek (T3i1); Replace the results of function EnglishUzbek (T3i1) to T2;

Here EnglishUzbek (T3i1) is the function written in [2] based on English-Uzbek translation direction algorithm. EnglishUzbek (T3i1) function is as following, accordingly used some signs to write function:

1. ET4k1 – Uzbek and English the structures that are suited each other 1≤ k1≤m2;
2. Load the function in the sentence of the word input T3i1 massive to E8k;
3. Find proper the structure sentence to E8k from ET4k1;
4. Take clause elements of the words found in ET4k1 and load to T2;

These tags represented in the following process:

```
private String  EngUzb(String          suz)     throws
ObjectNotFoundException {
int uzakId=0; String uzbSuz=””; int gapBulagiId=0;
UzakEnglish   ue=uzakSuzEnglishDao. getUzakEnglishByword(suz);
if(ue.getUzakEnglish().equals(suz)){ uzakId=ue.getUzakSuzlarId();
List<UzakSuzlar>      usList=uzakSuzUzbekDao. getuzakSuzlarListByRId(uzakId)
for (UzakSuzlar us : usList) { uzbSuz=us.getUzakSuzlar();}}else{
```

```
YasamaEnglish ye=yasamaSuzEnglishDao. getYasamaEnglishByWord(suz)
if(suz.equals(ye.getYasamaEnglish())){ uzakId=ye.getYasamaSuzlarId();
YasamaSuzlar yu=(YasamaSuzlar) yasamaSuzUzbekDao.getYasa
maSuzlarListByRId(uzakId);
uzbSuz=yu.getYasamaSuzlar();}else { uzbSuz=suz;}}
return uzbSuz;
}
```

In conclusion we may say that although our investigation on machine translation system seems a bit a simple, there are very pivotal issues should be done in terms of linguistic models. According to this rule based translation is important for non familiar and relative languages like English and Uzbek. In the future, our research will be directed multilingual machine translation system for the Uzbek language.

REFERENCES:

Single author Article:

1. Ахмедова Х.И. «Моделлашган компютер таржимаси технологиясининг алгоритмлари» Амалий математика ва информацион технологияларнинг долзарб муаммолари Ал-Хоразмий 2014. Самарқанд 15- 17-сентабр 2014 йил.
2. Ҳакимов М.Х. Математические модели узбекского языка. ЎзМУ хабарлари, № 3, 2010, с. 185–188.

Two/multiple authors Article:

1. Абдураҳмонова Н.& Ҳакимов М.Х. Логико-лингвистические модели слов ипредложений английского языка для многоязычных ситуаций компьютернеого перевода. 1-я Международная конференция “Компьютерная обработка тюркиских языков. Латинизация письменности” Казахстан, Астана, 2013, С. 302–306.
2. Абдурахмонова Н. & Ҳакимов М.Х. Семантические базы английского языка для многоязычной ситуации компьютерного перевода. Труды научной конференции «Проблемы современной математики» 22-23 апреля 2011 г., г. Карши, с. 311–314.

Chapter in a book:

1. Марчук Ю.Н. Компютерная лингвистика (учеб. пособ.) Москва, Восток Запад, 2007, 61–б.