

EXAMINING THE PROGRAM OF MACHINE TRANSLATION FROM KAZAKH TO CHINESE LANGUAGE BASED ON THE RULES OF KAZAKH LANGUAGE

Kaden Kenzhekhan, Doctorate of Information technologies department University of Xinjiang, Urumchi, China Kenjehan89@mail.ru

Gulila Altenbek, College of Information Science and Engineering, Xinjiang University, Urumqi, Xinjiang, 830046, China Xinjiang Laboratory of Multi-language Information Technology The Base of Kazakh and Kirghiz Language of National Language Resource Monitoring and Research Centre Minority Languages, gla@xju.edu.cn

Unzila Kamanur L.N. Gumilyov Eurasian National University, Astana, 010000, Kazakhstan Unzila.88@mail.ru

Machine translation is a type of translation from one language to another performed by a computer. In order to solve issues arisen in the area of machine translation it is significant to have a clear understanding of the context and rules of particular language. The presence of vocabulary with sufficient amount of words guarantees the quality of the translation results. Furthermore, the greatest difficulty encountered in this systems lies on creating this vocabulary. Nowadays, especially at the end of twenty first century the scope and area of translation have widened considerably. It can be noticed that many years before the concentration was on the machine translation from Chinese to Kazakh language. However, today due to the wide demand on worldwide exchange of information to be in a high quality translation from European and East languages to Kazakh language has become popular trend in all spheres of life. As a result, in order to create wide amount of vocabulary the documents from legal and social political areas have been translated. The science of linguistic translation of world's linguistic mainly copes with issues such as terms and term phrases in this sphere. According to the statistics today, there are more than 2000 labels in Chinese language related to the translation. However, in Kazakh language they are directly translated from Chinese language. The purpose of this paper is to perform good quality translation from Chinese to Kazakh language based on the Kazakh language and machine translation rules taken from the texts .

Keywords: Machine translation, Natural language

ИССЛЕДОВАНИЕ ПРОГРАММЫ МАШИННОГО ПЕРЕВОДА С КАЗАХСКОГО ЯЗЫКА НА КИТАЙСКИЙ ЯЗЫК, ОСНОВАННЫЙ НА ПРАВИЛАХ КАЗАХСКОГО ЯЗЫКА

Каден Кенжехан, докторант отдела «Информационных технологий» Университет Синьцзянь, Урумчи, Китай, Kenjehan89@mail.ru

Гулила Алтенбек, колледж информационной науки и разработки, Университет Синьцзянь, Урумчи, Синьцзянь, 830046, Китай, Синьцзянская лаборатория многоязычной информационной технологии, основа казахского и Киргизского языка, центр контроля ресурса национального языка и исследовательские языки меньшинства. gla@xju.edu.cn

Унзила Каманур Евразийский национальный университет им.Л.Н.Гумилева Астана, 010000 Казахстан Unzila.88@mail.ru

Машинный перевод - вид перевода с одного языка на другой выполняется компьютером. Для того, чтобы решить проблемы, появляющиеся в областях машинного перевода, существенно иметь ясное понимание контекста и правила специфического языка. Присутствие словаря с достаточным количеством слов гарантирует качество результатов перевода. Кроме того, самые большие трудности встречаются в этой системе на создании этого словаря. В настоящее время, особенно в конце двадцать первого столетия возможность

и область перевода расширилась значительно. Это может быть замечено, что много лет перед тем, как концентрация была на машинном переводе от китайского на казахский язык. Как-когда-либо, сегодня из-за широкого востребования на всемирном обмене информации, чтобы быть на высоком качественном переводе с Европейских и Восточных языков на Казахский, стало популярной тенденцией во всех областях жизни. В результате, для того, чтобы создать широкое количество словаря, документы от законных и социальных политических областей были переведены. Наука лингвистического перевода миров, лингвистических главным образом, покрывает с проблемами как например термины и фразами срока в этой сфере. Согласно статистике сегодня есть более чем 2000 этикеток на Китайском языке имеющие отношение к переводу. Но на Казахском языке они непосредственно переводятся от Китайского языка. Цель этой статьи - выполнять хороший качественный перевод от китайского к Казахский язык, основанному на Казахском языке и правилах машинного перевода, взятых из текстов .

Ключевые слова: Машинный перевод, Естественный язык

Introduction

Our purpose is to open new areas of machine programming in the context of translation from Chinese language to Kazakh language, especially we focus on defining the models of rules which will be particularly used in programming languages. More generally, does not matter what languages are used for translation activities it is required to fully change all the texts used in communication context so as the main context is understandable. As a consequence, all the translated texts are used for analysis of defining the unique rules used to translate from one to another particular language. It is noticeable that information and culture exchange between nations are taking place widely. Thus, the society and circumstances demand the systematic language among Turkish language nations which has been automatic in satisfactory level.

1. The importance of the research topic

Kazakh language uses two types of the writing style Arabic and Cirillic. In the Cirillic, the writing starts from the left of the page and continues to the right, whereas Arabic in opposite is known as RTL(right-to-left) language. Thus, Arabic being non similar to Chinese makes difficult of using in computer based translation since all writing stiles are one stylized for letter linkage. The link between words in Kazakh language is relatively complicated. We focus on making the translation from Cyrillic to Chinese in a high level based on the research and investigation, so that when information in Kazakh language is analysed it will be clear understandable.

2. Methods used in machine translation

According to the history of Machine translation from the early time divided as Traditional Machine Translation Method (TMTM) and Emerging Machine Translation Method, қысқаша (EMTM). TMTM is considered as being a complicated type of machine translation with rules and it has been covered in sufficient amounts of publication. This type of translation shows good results based on its rules and vocabulary. EMTM was developed based on TMTM and can be considered as Corpus Based Machine Translation (CBMT). CBMT is divided into two groups: Statistics Based and Example Based machine translation. Our systems mostly use rule-based machine translation system which works by conducting an analysis on sentences by using the rules.

3. The main working principals of machine translation

Our vocabulary consist of more than 20 000 verbs, nouns and phrases.

ID	HY	py	KZ	CX
1252	论文	lùn wén	мақала	n
1253	患者	huàn zhě	науқас	n
1254	热情	rè qíng	ықылас	n
1255	策略	cè lüè	тактика	n
1256	老人	lǎo rén	қарт	n
1257	通讯	tōng xùn	байланыс	n
1258	高中	gāo zhōng	орта мектеп	n
1259	同志	tóng zhì	жолдас	n
1260	昨日	zuó rì	күнi кеше	n
1261	外面	wài miàn	сырт	n
1262	部队	bù duì	аскери бөлiм	n
1263	大厦	dà shà	ғимарат	n
1264	红色	hóng sè	қызыл	n
1265	意外	yì wài	тұтқиыл	n
1266	大哥	dà gē	үлкен аға	n
1267	天堂	tiān táng	жұмақ	n
1268	兴趣	xìng qù	ынта	n

名称name	表格 form
名词 noun	n
动词 verb	v
形容词 <u>aj</u>	a
时间 time	t
代词 pronoun	p
成语 idiom	l
量词 quantifier	L
数词 numeral	num

Figure 1. Vocabulary

Kazakh language rules of cases has following prefixes added at the end of the words



Figure 2. Rules of Kazakh language

The performance of translation program

```
#region//如果hatan以结尾
if (U_Y_D(gu.kz) && (guo.hy.Equals("给") || guo.hy.Equals("往")))
{
    if (jwa > jin)
    {
        gu.kz = gu.kz + "ға";
    }
    else
    {
        gu.kz = gu.kz + "ге";
    }
}
#endregion
#region//如果hatan以结尾
else if (H_T(gu.kz) && (guo.hy.Equals("给") || guo.hy.Equals("往")))
{
    if (jwa > jin)
    {
        gu.kz = gu.kz + "қа";
    }
    else
    {
        gu.kz = gu.kz + "ке";
    }
}
}
```

Figure 3. Program for performing the translation

Results taken from the translation



Figure 4. Translation results
Conclusion

The experiments has shown that this system's performance of translation from Chinese to Kazakh language in satisfactory level. In future, it is planned to improve the translation of verb tenses as well as use of the plural nouns, since compared with English language their rules are much more and complicated depending on the vowels and consonants.

Defining the tenses is also considered as a complex process , because it is necessary to perform analysis. Afterwards using taken analysis the program synthesizer has to be created.

Creating the synthesizer include several difficulties since the results needs to be matched with the tense used in original text. Otherwise the sentence meaning can be modified and making the translator confused. All of this abovementioned cases is fully linked to the distinct features of Kazakh language compared with other languages.

References

1. Gulila Altenbek. and Dawel,A. and Muheyat,N. 2009. A Study of Word Tagging Corpus for the Modern Kazakh Language, Journal of Xinjiang University, 26(4):323–326
2. <http://check.cnki.net/user/>,

UDC 004.8

SPOKEN TERM DETECTION FOR KAZAKH LANGUAGE

Zhanibek Kozhirkbayev National Laboratory Astana,Nazarbayev University, Kazakhstan e-mail: zhanibek.kozhirkbayev@nu.edu.kz

Muslima Karabalayeva National Laboratory Astana,Nazarbayev University,Kazakhstan, e-mail: zhyessenbayev@nu.edu.kz

Zhandos Yessenbayev National Laboratory Astana,Nazarbayev University, Kazakhstan e-mail: muslima.karabalayeva@nu.edu.kz

The paper presents a spoken term detection system for Kazakh language in which significant improvements are obtained through modifying speech-to-text process used for generating word-based lattices. These lattices are indexed and used for the keyword search later. Spoken Term Detection systems quickly discover the occurrence of a term, which might be just a word or sequence of words, in a large audio set of heterogeneous speech records. The paper provides an overview of a speech-to-text and keyword search system architecture built primarily on the top of the Kaldi toolkit and expands on a few highlights. Our aim was to develop a general system pipeline which could be advanced regarding phonological and linguistic features of Kazakh language in order to detect OOV keywords.

Keywords: Speech Retrieval, Lattice Indexing, Spoken Term Detection, Speech Recognition, Keyword Search

ПОИСК РАЗГОВОРНОГО ТЕРМИНА НА КАЗАХСКОМ ЯЗЫКЕ

Кожирбаев Жанибек Мамбеткаримович Национальная лаборатория Астаны, Назарбаев Университет, Казахстан e-mail: zhanibek.kozhirkbayev@nu.edu.kz

Карабалаева Муслима Хизиятовна Национальная лаборатория Астаны, Назарбаев Университет, Казахстан e-mail: muslima.karabalayeva@nu.edu.kz

Есенбаев Жандос Аманбаевич Национальная лаборатория Астаны, Назарбаев Университет, Казахстан e-mail: zhyessenbayev@nu.edu.kz

В статье представлена система для обнаружения разговорного термина на казахском языке, в котором получены значительные улучшения путем модификации процесса речи в текстовый формат, используемый для создания слов на структурной (lattice) основе. Эта