Research on Computer Translation Software Technology for Network Client

Zheng Zhao[1, a]

[1] Shandong University, Weihai, Shandong, 264209

[a] email

Keywords: Computer Translation Software, Network Client Orientation, Translation Technology

Abstract. With the development of society and the fast pace of technology, computer-aided translation technology has been used as an important adjunct widely used in software localization and technical translation work. This reality needs to also stimulate the translation of teaching the importance of computer-assisted translation tools, and started a number of useful attempt. Thus, training tools and language translations service management will become increasingly important day. However, with many new things, computer-assisted translation technologies Course at the initial stages, either teaching or practice of instructional design is facing many difficulties.

Introduction

Machine translation process generally can be divided into five parts) input source language, source language analysis, the source language to the target language conversion target language generation, target language output. It can be divided into rule-based and corpus into two categories. Among them, the corpus-based approach can also be divided based on statistical methods and instance-based. Rule-based machine translation technology is the most mature, and it is by far the most widely used, currently affecting machine translation systems are rule-based. We are translated using the corpus as a source of knowledge based on statistical machine translation methods and instance-based. The difference is that: the method based on statistics, get the translation of knowledge before the translation is completed, the translation process is no longer used corpus; and in the method based on the examples, but also in the translation process queries and use Corpus.

Based Machine Translation Thought instance was first used by the famous Japanese machine translation expert Makoto Nagao raised. The basic idea is not through deep analysis, but only through the existing experience and knowledge, translated by analogy principle. In future research in machine translation, learn from each other a variety of methods, convergence trend will become increasingly evident. Rule-based approach and corpus-based approach combining machine translation and translation memory combination, is likely to be the mainstream of the future direction of research and development.

Development of Translation Technology to Provide A Real-World Applications

At least in the perspective of innovation, science and technology did not let us down. After decades of development, translation technology, we now have two Keziliyong translation technology system. One is an early start, long-time machine translation (Machine Translation, MT), one is started late, but the rapid development of computer-aided translation technology (Computer Assisted Translation, CAT). Below we will briefly what are their development and break both aspects.

Corpus-based statistical translation system and booming, suddenly let people shines, it is currently offering online machine translation services agency or organization generally accepted. Such translation systems based on parallel corpus collected, through corpus analysis of translation match. Thus, the corpus collected for this type of machine translation systems is essential. At present, several more famous online translation service provider also has its own set of search engine technology. Such as Google Translator, Bing Translator and domestic NetEase's Youdao online translation. We can see three common characteristic is their translation services are established in the corresponding search engine based on the (Google, Bing and proper way). The

vast "network of sea" in excellent search technology also laid the basis for their doing well.

Compared to machine translation, computer-assisted translation is more realistic, some of the more extensive range of applications. The basic principle is to work, doing the translation software translators will be memorized (translation memory), so the next translator encountered similar translation problems, the software will automatically match the case of previous translations, translator for selection . In addition, the translator can also establish a glossary to standardize the terminology of the document, so that the full text of the terms consistent.

Although computer-assisted translation process of manual intervention or dominant part, but we see that this is the current computer-aided tools for translation reality. This seemingly backward move would effectively solve practical problems is undoubtedly a big step forward. Now the main computer-aided translation software German Trados (Trados) software (SDL has been that SDL acquired), it is now the mainstream software industry, in addition to the SDL (SDL) of SDLX (with Trados has a), Atril's Déjà Vu. Three industry known as the Big Three, there are three pillars. In addition, domestic and snowman Masanobu CAT CAT also have different characteristics. However, in practice, computer-assisted translation and machine translation is now no longer two things are diametrically opposed, but gradually towards integration and mutual support. For example, Google Translator Toolkit Translation underneath, it is difficult to define a machine translation tool or a computer-aided translation tools. This integration is undoubtedly more realistic needs, but also to broaden the application field of translation technology. For translators, the translation technology becomes easier to use, regardless of function or quality has been further improved. In short, for all parties it is regarded as good news.

Key Technology of Computer Aided Translation Memory

Translation memory technology can help us to solve many practical problems targeted, especially as machinery manufacturing process documents in the statement, you can reduce the amount of duplication, harmonization of terminology for translators style converge. But in the field of machinery manufacturing it has never materialized, and some key technologies involved in the translation memory has yet to continue to study. Actually, using computers to process knowledge, the main problem is the three circles: First, get knowledge; the second is knowledge; the third is to use knowledge, then to discuss some of the key research techniques, is for these three issues raised solutions to solve these key technologies will translate the whole process smoother to minimize bottlenecks, improve translation speed and translation quality.

Important by the following number of key technologies based translation memory technology: (1) enter the Chinese to be translated sentences, word segmentation based on a bilingual dictionary, to get the correct word segmentation form; (2) to certain characteristics to be translated sentence as a search query Examples of the library to retrieve similar or the same sentence, in accordance with the size and similarity to select the most similar sentence, that similarity calculation; (3) based on certain theoretical alignment, the alignment be translated sentences and sentence library sentences, and align sentence library Chinese words and sentences in English sentences words; (4) the use of certain English generative theory, generate translations to meet the requirements; (5) required the whole process of translation memory, tailor-made translation memory associated libraries.

Written expression based on Chinese characters as the smallest unit, but in natural language understanding among the smallest meaningful word processing unit. The division did not sign that there is no word boundary character string conversion to comply with the actual language of the word string that is established in the written word in the Chinese border, which is the Chinese word task. Chinese natural language translation system outside a Chinese understanding, the first encounter is insurmountable Chinese automatic segmentation phase. Use between the word and the word co-occurrence frequencies between words as between word basis, you can not build a good advance word dictionary in the statistics-based approach. The advantage of this method is that it is not to limit the field of application and is not limited to the previously established sub-word dictionary. Based on statistical methods require large-scale training texts for training the model parameters, and whether it is training or the actual text segmentation, generally requires a large

amount of computation [3 ,, in this translation system should not be used.

Examples of the library there is a lot already translated sentence can be used as a reference template for later translation. To select a similar sentence, must determine what sentence case library can be used to translate the sentence to be translated input, which must be translated to determine sentences and sentence similarity (hereinafter referred to as the similarity), it is appropriate to establish a similarity criteria. Comparison on vocabulary level be translated sentences and sentence. We should take the most similar sentence translation as a template to deal with the translation sentence translated, you must first select the most similar sentence out. According to the overall structure similar to the initial sentence retrieved after comparison be translated sentences and sentence on the similarities and differences in the use of vocabulary, mainly relatively speech, semantics, and given the similarity of two sentences on the words, sentences, and we have to be translated Examples of the degree of similarity in vocabulary level called trimming similarity. Finally, the overall similarity and related trim similarity weighted summation, accurate similarity to be translated sentences and between sentences.

Bilingual alignment is an important issue in natural language processing research. Corpus establishing correspondence between the original and the translation between the same grammatical units, called alignment. Bilingual text align relationship exists multiple levels, including the level of alignment of the paragraph, the sentence level, phrase level, word level.

By similarity calculation, already we know to be translated sentence and the example sentence selected on the sentence level are similar, can be seen as a sentence-level alignment. Align Technology paper includes translation between Chinese sentences and sentences in words, and words are aligned alignment fragments on the exchange level, refer to the alignment word level. Align be translated sentences and sentences between Chinese mainly through the way the similarity calculation obtained, in particular fine-tuning of calculating the similarity is substantially complete alignment between words and sentences to be translated Chinese sentence. Chinese example sentence and a translation sentence sentence previously been stored in the library, you can use directly extracted. They are aligned in this paper is mainly by means of human-computer interaction in the post-editing stage approach to achieve. Fragment alignment comparison relatively trouble, with the aid of a sub-block fragments stored in the library as an intermediate state is compared to Semantic Extraction center, and then compare the central meaning of the words are the same two fragments to determine whether the fragment is aligned.

After the similarity calculation and alignment technology, translation entire generation is an important part of the translation work, the computer automatically work is done by the last step, how a similar sentence translation as a template to guide the translation of the sentence to be generated is generated translation to be studied Content. According to the sentence to be translated and the translation sentence align relationship alignment phase completed to constitute a pattern sentence translation as a guide, according to the sentence needs to be translated to do some conversion operations such as copy and replace operation to complete the translation work is generated.

Conclusion

This paper focuses on the major problems of translation memory technology to solve launched a more in-depth study, including similarity calculation, alignment method, text generation and post-editing and other issues. Text adopted to solve the problem is different from the problem-solving ideas, try to avoid some of the calculation results obtained by the probability approach, especially in the calculation of alignment, text generation and word choice, etc., the theory set forth herein in a practical, simple, easy implementation of the principle of the study.

References

[1] Brace C. Language Automation at the European Commission, in Sprung, RC (. Ed) Translating into Success: Cutting-Edge Strategies for Going Multilingual in a Global Age [M]. Amsterdam

and Philadelphia:. John Benjamins, 2000, 219-24.

[2] Chesterman, Andrew, Wagner E.Can Theory Help Translators [M]. Manchester: St. Jerome, 2002.

[3] Huang Heyan, et al The design and implementation principle of an interactive hybrid strategies machine translation system:.. IHSMTS [J]. The Int'l Conf in Machine Translation & Computer Language Information Processing, Beijing, 1999.

[4] Chen, J., R. Chau, and CH. Yeh. Discovering Parallel Text from the World Wide Web. In Proceedings of the second workshop on Australasian Information Security [J]. Data Mining and Web Intelligence and Software Internationalization, 1991, 1-2.

[5] Nie, JY, MSP Isabelle, and R. Durand. Cross-language Information Retrieval based on Parallel Texts and Automatic Mining of Parallel Texts from the Web [C]. In Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development, 1999.