

Multi-user and multi-language collaborative speech translation system and method in the era of big data

Mengyang Xing

National University of Defense Technology

15838861121@163.com

Abstract. In the rapid development of information technology and network technology, the cultural knowledge of human society blends with each other, and people have more and more ways to obtain information. Faced with the continuous growth and various types of information composed of different languages, researchers have proposed to use machine translation technology to build auxiliary translation system, which can play an important role in daily learning and life work. After understanding the research and application status of speech translation system in the era of big data, this paper mainly explores the architecture of multi-user and multi-language collaborative speech translation system according to the application characteristics of multi-language assisted translation system in the new era, and puts forward practical application methods from the perspective of long-term development of the industry, so as to let the world economy and culture communicate with each other. Fully demonstrate the translation efficiency and application quality of the design system.

Keywords: Big data; Multi-user; Multilingual; Synergy; Speech translation.

1. Introduction

In the current information age, the demand for translation shows an increasing trend year by year. Since the effect of machine translation is difficult to meet the practical needs, how to improve the quality of machine translation and ensure the efficiency of practical work has become the main issue focused on the field of machine translation. Machine translation is the basic component of artificial intelligence. How to understand human language is the focus of research and development of intelligent technology. Human beings are more suitable for abstract thinking, while computers have advantages in mass storage and rapid retrieval, calculation and other aspects. Integrating humans and computers together and making full use of their own basic characteristics has become the main direction of the development of artificial intelligence in the new era. The study of translation technology in China started relatively late, but with China playing an increasingly important role in the process of globalization in recent years, there is a strong demand for the development of translation technology. After learning from the development experience of the same industry abroad, China's translation industry has made certain achievements in the field of translation technology from 2007 to 2008. Although there are still some problems in urgent need of improvement, this will inevitably have a far-reaching impact on China's translation industry. [1-6] Nowadays, China's science and technology departments and financial departments have provided special funds as support while strengthening the humidity in the research and development of translation technology. Chinese and multilingual processing software and Chinese translation software with advanced linguistic theory as the core are listed in the 2008 Technological Innovation funding support plan for smes. At the 18th World Translation Congress in 2008, the China Agreement set aside a special forum on translation tools, terminology management and translation standards, thus promoting the development of translation technology in China.

Dictionaries are widely used as technical tools, both language learners and professional translators will come into contact with and master this content. Since the early stand-alone version

has developed, different product types have emerged in domestic and foreign markets, such as mobile dictionaries, online dictionaries, electronic dictionaries and so on. Nowadays, some scholars use user-published vocabulary design dictionary compilation system in their research, and regard dictionaries and vocabulary published by major publishing houses as the main data sources, so as to facilitate Internet users to query and learn. Since there are many defects in the translation methods of traditional published dictionaries, especially after the emergence of Internet platforms, users have higher and higher requirements for dictionary translation, so how to use modern science and technology to develop multi-user and multi-language collaborative speech translation system is the main topic of comprehensive discussion among scientists. After understanding the development status of speech translation system and related technologies in the era of big data, this paper mainly explores the design and implementation of multi-user and multi-language collaborative speech translation system in the era of big data, so as to provide technical support for the innovation and development of translation industry in the new era.

2. Method

2.1 Multi-User Collaboration

Interactive translation is the main topic of human-machine integration technology research and development. In the era of big data, communication between people is becoming more and more convenient, which makes collaborative work between people possible. In order to further improve the efficiency and quality of translation and meet the interactive information needs of all users, a collaborative mechanism can be used to build a translation system as shown in Figure 1 below:[7-9]

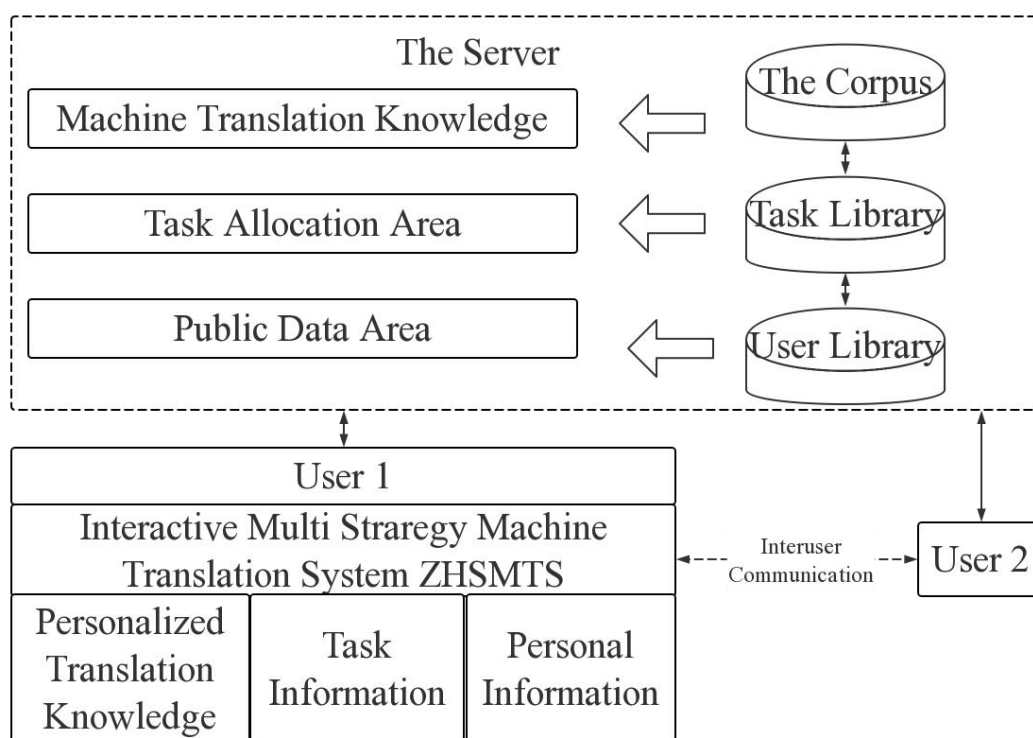


Figure 1 Translation system

The overall system design adopts the client-server approach, which includes two parts: client and server. The server side mainly stores users' public translation knowledge, translation tasks and information of translators, which can provide users with areas for collaborative work and task

allocation, and at the same time, it has the function of automatically extracting translation knowledge from corpus and self-learning. The client will store the user's personalized translation knowledge, task information and personal information.

2.2 Multilingual assisted translation system

Assisted translation refers to a translation method that uses computer information processing power and manual intervention to transform the source language text into the target language. It will make use of the automatic team and the language text to carry out grammatical and lexical analysis before the translation, search and query in different databases, and finally check and correct the target language translation output by automatic translation after statistical analysis and optimization processing. In this way, you can get a translated text that meets the needs of the user. Based on the analysis of the assisted translation flow chart shown in Figure 2 below, it can be seen that manual intervention mainly appears in the translation stage. On the one hand, it is used to provide text with clear semantics. When selecting the text for machine translation, it is necessary to provide phrases or sentences with complete semantics and rules as much as possible to ensure the accuracy of translation. On the other hand, the machine translation results will be appropriately modified and manually proofread to achieve the desired set of satisfaction. While meeting fundamental needs, the multilingual assisted translation system will use a variety of strategies and unique technologies to improve the accuracy and effectiveness of translation, so as to meet the language translation needs of multiple users.[10-13]

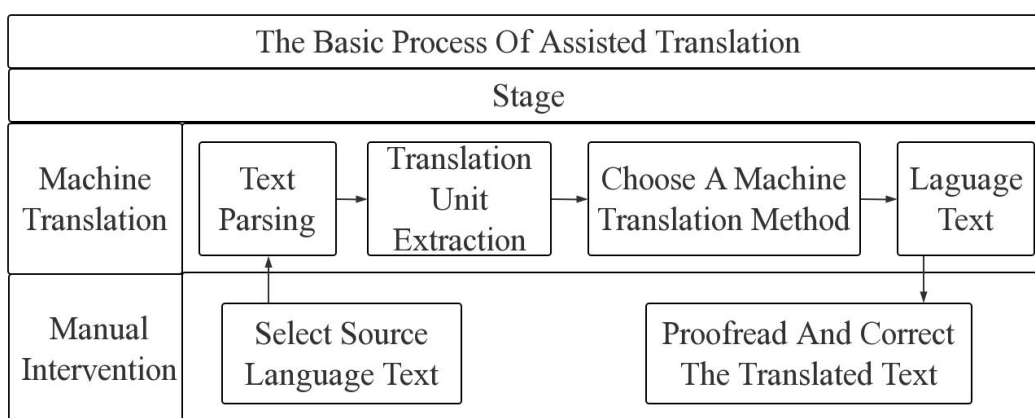


Figure 2 Flow chart of auxiliary translation

2.3 Overall Architecture

The multi-user and multi-language assisted translation system can choose two modes: stand-alone version and online version. The former only contains core function module, local interface module and local file storage module, while the latter adds several modules such as user online function, server interface, core function, application service, database interface and background database storage. From the perspective of practical application, the stand-alone system architecture has two functions, on the one hand refers to the machine translation service, on the other hand refers to the file reading and writing function; The network version system architecture can use the protocol and server port communication, call the auxiliary interface provided by the server, and effectively call the Web Service mode. The translation work of the online version system can provide project management services, while the stand-alone version system can only provide services for personal translation, the specific price is shown in Figure 3 below:

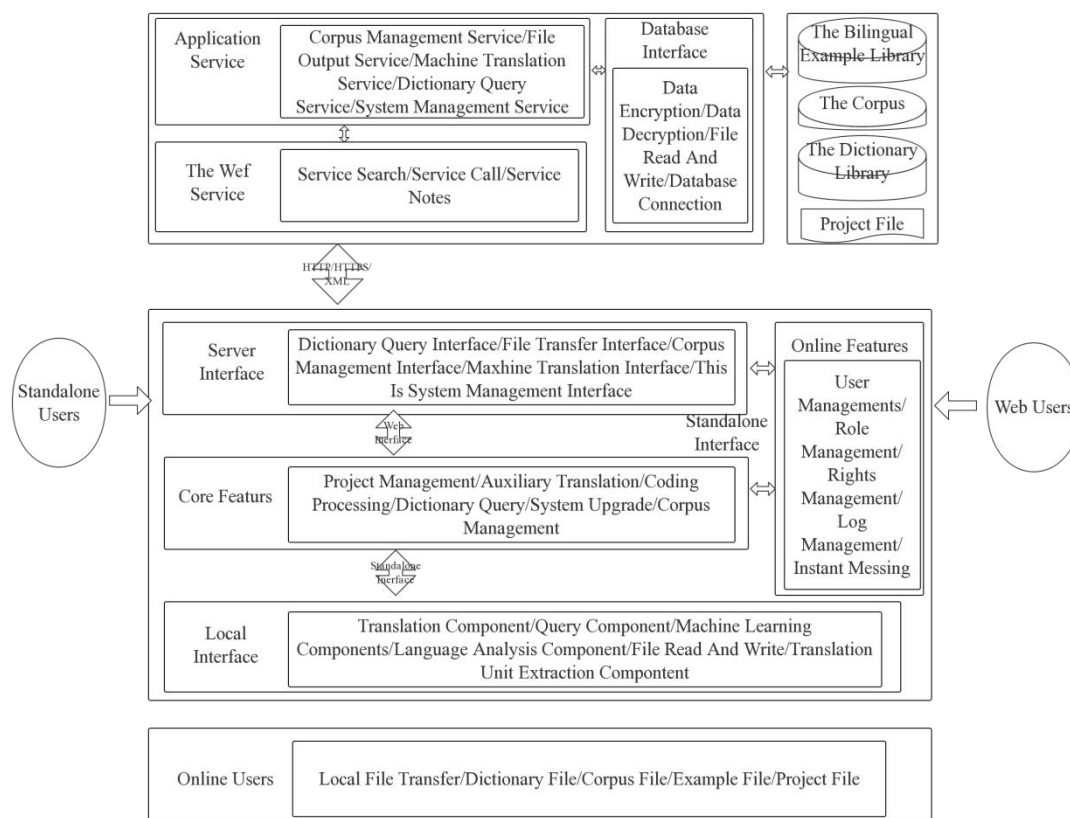


Figure 3 System architecture diagram

Based on the above analysis, we can see that auxiliary translation application services include the following points: First, machine translation services, mainly for different types of systems, users to provide source language text words, analysis and grammar analysis and other service functions, is an important module of the overall system operation. The system can usually provide three kinds of machine translation services, the first is strength-based translation, the second is statistics-based translation, and the last is corpus-based translation. Users can choose the translation method according to the characteristics of the translated text and the nature of the work, so as to ensure the effectiveness of the translation work. Second, corpus management services can provide retrieval, inquiry, protection and other services for the systematic corpus, which is the basic basis of machine translation. Effective services can facilitate users to obtain high-quality corpus. Third, dictionary query service, mainly for translators to provide word and phrase query services; Fourth, system management services, mainly for system users to provide basic services such as project management, personnel management; Fifth, file transfer services, mainly provide users with basic services such as file upload, download, encryption and decryption in project management.

Based on the analysis of the system flow chart shown in FIG. 4 below, it can be seen that the translation project of the multi-user and multi-language collaborative speech translation system includes steps such as task assignment, personnel assignment, schedule control and work review. The task will be decomposed by means of splitting and merging according to the oversized text, and the XML format file will be used for annotation, and the translation service will be completed in an orderly manner.[14-15]

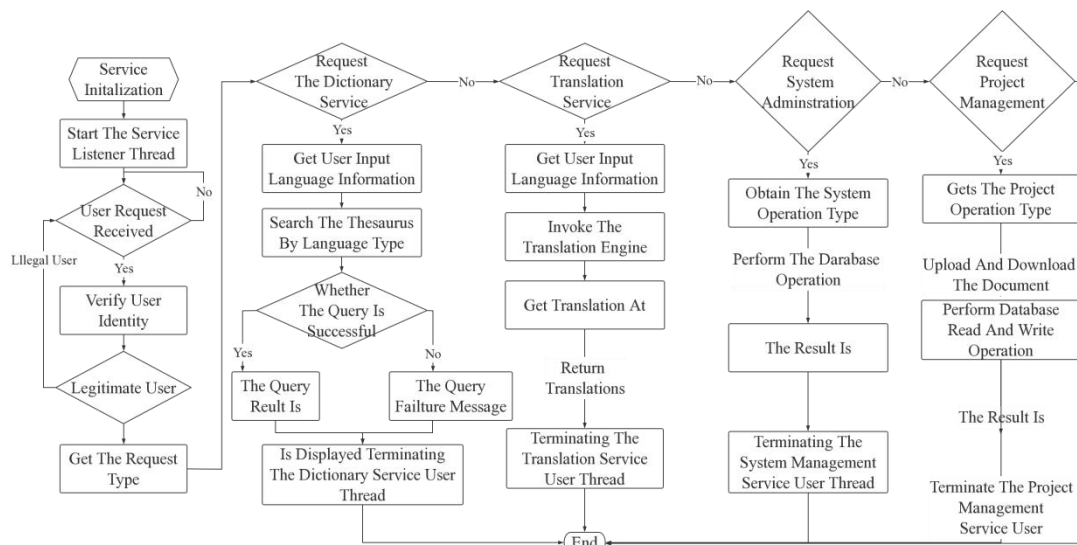


Figure 4 Flowchart of the system

3. Result analysis

Interactive translation, as the basic content of human-machine combination research, although the interaction ability of individuals and individual computers is limited, and the real real-time knowledge sharing cannot be realized, with the continuous development of social economy and science and technology, the collaborative work form between people has become possible, and modern science and technology has been used to build a new voice translation system architecture. Fully demonstrate the computing power and network functions of the computer, according to the needs of users in various industries, put forward strong performance and good effect of technical tools. Interactive multi-strategy machine translation system is a new translation system built by integrating multi-knowledge description, experience memory, heuristic reasoning, statistical knowledge, etc. Compared with the traditional single-machine interactive translation mode, it can improve the efficiency and quality of the system operation, and meet the needs of multi-user interactive translation in the new era. From the perspective of practical application, the types of translation operations of users are shown in Table 1 below:

Table 1 Analysis of user translation operation types

serial number	Operation description	Related corpus
1	Make all changes to the sentence translation.	translation
2	Add a paragraph (word or phrase) to the translated sentence.	translation
3	Modify the meaning paragraph in the translated sentence	translation
4	Delete the meaning paragraph in the translated sentence	translation
5	Adjust the order of Italian paragraphs in the translated sentence	translation
6	Adjust the corresponding relationship of words	Correspondence information
7	Add new words and phrases	word stock
8	Sentence-breaking the original text	original text
9	Merge sentences on the original text	original text

Based on the analysis of the above table, it can be seen that the system adopts the multi-user mode, and user information plays an important role in the overall system operation, including user permissions, passwords and names. During the operation of the translation system, the unique

phenomena and grammar in translation can be directly stored, which can be regarded as the knowledge resources of the translation system.

On the basis of understanding the research and development status of speech translation systems and related technologies in the era of big data, this paper identifies the existing problems in the application of the existing assisted translation system, integrates the functional characteristics and solutions of the multilingual assisted translation system, and finally provides effective technical means for multi-user translation services in the new era. The multi-user and multi-language auxiliary translation system can choose two modes of interactive translation and automatic translation, which can provide corresponding operation interfaces according to different types of language requirements, and fully support the writing habits and display operations of users in various industries. In the future, China should continue to explore the framework of multi-user and multi-language translation system in the new era and learn from the unique advantages of existing translation systems at home and abroad. Only in this way can we lay the foundation for the technical development of our translation industry.

Conclusion

To sum up, the development of science and technology has led to the rise of computer and other industries, and the multi-user and multi-language collaborative speech translation system has also been comprehensively promoted. Although there are still many problems in existing research and application, with the continuous innovation of social, economic and technological means, it is inevitable that advanced knowledge and rich experience can be used in the future to create a more high-quality speech translation system architecture, change the relatively single way of use, and add more translation meanings and grammatical information. Truly meet the voice translation needs of multiple users in multiple languages in different fields. Therefore, at present, while increasing the theoretical research on speech translation technology, China has always strongly supported the application of multilingual translation systems, paid attention to the training and development of relevant professional and technical talents, and will formulate preferential policies from the long-term perspective of the industry, which will not only adapt to the development trend of economic globalization more quickly, but also provide technical support for cultural exchanges between countries. It can also enrich people's access to information and lay a solid foundation for cultural transmission between people

Reference

- [1] Libo Feng, GeFEI Feng, Weidong Zhan, et al. Current situation and prospect of computational linguistics in China [J]. *Language Science*, 2021, 020(005):491-499. (in Chinese)
- [2] Wen Liu. A method to improve the checking efficiency of multilingual terminology database by using Web crawler technology [J]. *Science and Technology Information*, 2023, 21(8):37-43.
- [3] Jing-cheng Cao, Xiao-Qian Wu, Qian Wang, et al. Research on parallel corpus construction for multi-language patent machine Translation [J]. *China Invention and Patents*, 2022(006):019.
- [4] There are hundreds of flowers in full bloom. Design of automatic English and Sign Language translation System based on Speech and Image fusion platform [J]. *China New Technology and New Products*, 2022(18):13-16.

- [5] Xiaodong Yang, Tian Tian, Jiaqi Wang, et al. A certificateless multi-user multi-keyword ciphertext retrieval scheme based on cloud edge collaboration [J]. *Journal of Communications*, 2022, 43(5):11.
- [6] Li Si, Jing Zhou. Cross-language retrieval function analysis and development strategy of "One Belt and One Road" multilingual shared database [J]. *Library and Information Work*, 2021, 65(3):8.
- [7] Yanxin Chen, Jia Dong. Study on the improvement of multilingual translation supply capacity in Yunnan Free Trade Zone under the background of RCEP [J]. *Comparative Research on Cultural Innovation*, 2022, 6(13):4.
- [8] Dong Yu, Wanying Xie, Shuhao Gu, et al. Multilingual Neural machine translation based on language association course learning [J]. *Computer Science*, 2022, 49(1):7.
- [9] Yuan Ping. Multilingual automatic translation telephone [J]. *Invention and Innovation: Big Science and Technology*, 2007(4):1.
- [10] Zheng Wang. Analysis of Multilingual intelligent Translation Service System [J]. *Journal of Integrated Circuit Applications*, 2019, 036(009):58-59.
- [11] Xiaojun Liu. Machine Aided Translation System based on Multilingual parallel corpus [D]. Beijing Institute of Information Control, 2008.
- [12] Jun Teng. China National Language Translation Bureau National Language intelligent speech translation software Conference and gift Ceremony held in Beijing [J]. *Journal of Ethnic Translation*, 2020(1):1.
- [13] Limin Du. The high-tech dawn of the new century: Speech Translation technology [J]. *Electronic World*, 2001(4):1.
- [14] Zheng Wang. Analysis of Multilingual intelligent Translation Service system [J]. *Integrated Circuit Applications*, 2019, 36(9):2.
- [15] Hongchen Jiang. Research on Multilingual Language recognition technology [D]. Institute of Automation, Chinese Academy of Sciences, 2007.