

An Interview of Cloud Translation Platform

Xiao-Yan Ren

Abstract—Cloud Translation platform is one of the main research issues in the field of computer-aided translation field. This paper reviews the domestic research status of cloud translation platform. Firstly, the paper presents a survey on the compute-aided translation and its application trend, describes the system architecture and workflow of the cloud translation system, and discusses the relations and differences between traditional CAT and the cloud translation platform. Then focused on the three key problems of aiding translation generation, post-editing and feedback learning, this paper analyzes the difficulties and key technology of cloud translation, and shows the domestic academic achievements and commercial application of the cloud translation platform. At last the paper outlook the prospects for future research of cloud translation platform.

Index Terms—Cloud translation platform, cloud computing, computer-aided translation.

I. INTRODUCTION

With the rapidly popular of globalization process, the language obstacle problems between different races and nations are increasingly obvious. Meanwhile the translation industry got developed quickly. In order to help translators work more efficiently and make full use of the computer's computing and storage capacity, the experts and scholars have carried out the research on computer-aided translation for more than fifty years. Traditional CAT techniques have encountered considerable difficulties when applied to translators. For example, if multiple spatially distributed interpreters need to complete a translation task collaboratively, the CAT software cannot dissolve such problems project schedule management, terminology consistency, personalized auxiliary translation and so on.

At present, most translators are depending more and more on the CAT software, such as SDL Trados, D     Vu X, YaXin CAT, YiJin Transwhiz and so on. The above CAT commercial systems are all composed of translation memory modular, terminology management system, alignment tools and translation project management, and other key technologies. Although the above commercial systems are full-featured, most CAT tools do not provide translation memory resource libraries when users buy the software. Even if some software companies provide the parallel corpus resources, the quality of resources and the quantity of the parallel corpus are both poor. A newly CAT software with poor translation-memory system will not have good performance. So the translators have to reconstruct the parallel corpus after buying the software, and they often

complain about such as the lack of accurate resources, the poor accuracy, labor cost and so on. And statistics show that China's translation industry is still in a relatively backward state, and half of the companies are composed of fewer than 10 employees. Most domestic translation companies are small, scattered and chaotic, and they also faced with insufficient reserves of translators, and the parallel corpus built fragmentally. So the current translation mode is still human-based translation, that translators make use of a variety of computerized tools to help them complete translation tasks and increase their productivity [1]. Meanwhile the domestic translation service is inefficiency and high cost, and high quality translation service in China is facing a huge market gap. To address these problems, researchers and some translation companies have proposed cloud translation platform, combining cloud platform and language processing technology. In this paper, the author reviewed the key technologies and the application of the cloud translation platform.

The rest of the paper is organized as follows. Section II firstly discussed the relations and differences between traditional CAT and the cloud translation, then described the cloud translation platform and its key technologies that are aiding translation generation, post-editing and feedback learning. Section III showed the domestic achievements and application of the cloud translation platform. Section IV concluded this paper and forecasted the prospects for future research of cloud translation platform.

II. CLOUD TRANSLATION PLATFORM AND ITS KEY TECHNOLOGIES

Cloud computing is considered to be the fourth IT industrial revolution after the large computer, personal computer, and the internet. The word "cloud" is a metaphor for network, Internet. Cloud computing platform provides services based "cloud" for developers to create applications. The developers can rely on the cloud platform to create a new SaaS (Software as a service) application. With the rapid development of the cloud computing industry, cloud computing has been increasingly applied in many fields. Therefore, the combination of traditional CAT and cloud computing has become a trend in the translation industry, thus the birth of cloud computing.

Cloud translation platform is collaborative translation platform based on the cloud platform, applies advanced computer and language processing technology to translation sector, and can realize the collaborative translation functions, organizing multiple spatially distributed translators to complete one translation task. Relying on cloud computing technology, cloud translation platform provides the public and translation industry with broad, safe and reliable Internet language translation service, which can improve the

Manuscript received June 30, 2015; revised October 12, 2015.

The author is with the college of information technology and computer science, China Three Gorges University, Yichang City, Hubei Province, China (e-mail: 50706780@qq.com).

efficiency of the whole translation industry and will have a bright future.

A. The System Architecture

The cloud translation platform is a new type of artificial online real-time translation system, which is composed of the cloud user and the cloud server. The system architecture of the cloud translation system is shown in Fig. 1.

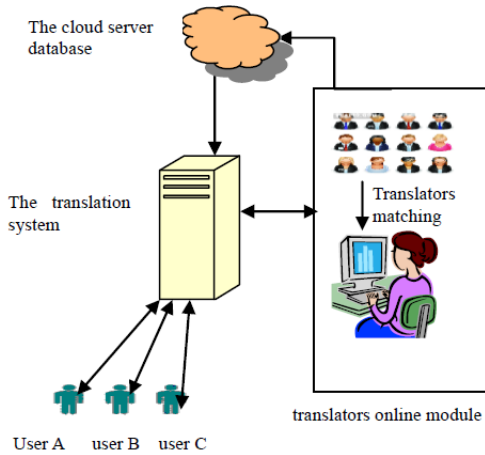


Fig. 1. The cloud translation platform's system architecture.

The cloud-users terminal includes cloud-users input module, cloud end-users transfer module and cloud-users output module. The cloud users' input module is used for the input of the cloud user's translation request, including the original text and the choice of translation target language. The cloud end-user transfer module transmits the cloud user's translation request to the cloud server through the communication network, and receives the translation information from the cloud server. The cloud end-user output module is used to output the feedback of cloud server's translation information.

The cloud server includes cloud server database, cloud server standard database, cloud server temporary database, cloud server transfer module, cloud server search module, cloud server update module. The cloud server database stores translation retrieval application procedure, seeks the artificial cloud service end help, saves the multilingual characters, the word, the sentence original text and the corresponding translation material. The cloud server standard database establishes and stores a cloud server database by manual input of all words, sentences and the corresponding translation data. The cloud server temporary database stores users' input in cloud server database, which are not the same as the target text link artificial cloud and cloud client requests translation server translation. The cloud server transport module receives the translation request from the user side, and transmits it to the cloud server and receives the artificial cloud server and transmits it to the end-users. The cloud server searching- module's main function is to retrieve the translation of the original text from the server database. The server update module transfers the words, sentences stored in the database of the cloud server temporary database, and the database of the server is updated by the server update module.

B. Work flow of the Cloud Translation System

As illustrated in Fig. 2, in the cloud translation project,

the translation of the work process is generally divided into three steps:

- 1) Translators conduct a preliminary translation. After the project started, the translators began to work, in order to achieve the desired quality objectives; the translators can get the support of two aspects, which are the software technical support and the quality control personnel's experience support. Translators can get some technical support through the cloud translation platform to improve the quality of the translated text. For example, the system will give a reference to the previous version of the translation, to avoid duplication of work and to improve the translation efficiency if the similar sentences that have been translated before the project. If the interpreters' translation do not match the terminology translation provided by customers built on the cloud server platform and provide customers the translation platform, built into translation, the system will give translators a hint and need interpreter's confirmation about the correctness of the translation, to avoid mistranslation. Similarly, if the number of the translated text is not consistent with the original text, the system will also remind the interpreters to confirm again. Therefore, translators will improve the quality of translation and translation efficiency; avoid misinterpretation and mistranslation with the help of the cloud platform. The work of an interpreter in a cloud translation platform is stored in the cloud server. Within the project staff, managers can login the translation platform, and easily obtain interpreters' translated text. In the translation project, experienced supervision of quality control staff interpreters can be arranged among the translation process, detect quality problems in the process of translation, and feedback to the interpreter. So translators can ensure the quality of translated text at minimum cost to correct quality problems under the online quality control personnel supervision.
- 2) The reviewers carry on the editing work. Second-Person Review is a common recognition in the translation industry. Reviewers also work on the cloud translation platform, as well as the translators to obtain the system's technical support. In addition, if reviewers modify a sentence, then all similar sentences in the project will be automatically corrected. The cloud system may automatically distinguish the script text edited by the reviewers through different color discrimination, and refer automatically to interpreters.
- 3) To revise the translation according to customer's feedback. In all cloud translation projects, the project results will be submitted to the customers after the reviewing work is completed, but it does not mean the end of the project. Usually the customers evaluate the translation and give revised opinions based on their quality goals and quality expectations. The translators modify the translated texts as far as possible according to the customers' requirements. In this stage, interpreters do the revision work on the platform, at the same time new terms are added to the terminology storehouse. The translation accepted by the customers will added to the translation memory and have the highest priority, namely if the translators in the future

work fall through the same or similar sentences, the system will choose the translated texts accepted by the customers for the interpreters.

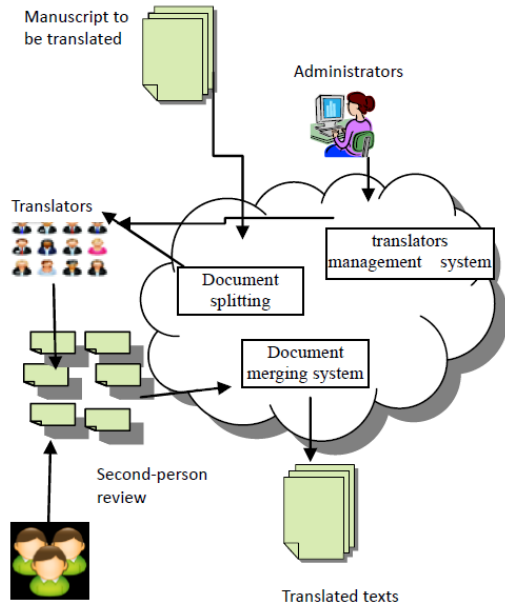


Fig. 2. The workflow of cloud translation system.

C. Comparison Between Traditional CAT and Cloud Translation

In the era of Web2.0, traditional CAT and cloud translation can be regarded as the two kind of main information technology translation patterns. Compared with the existing machine translation software, translation system has the characteristics of high accuracy, high efficiency of translation. Table I shows some differences between the traditional CAT and cloud translation through translation technology and translation environment.

TABLE I: THE COMPARISON BETWEEN TRADITIONAL CAT AND CLOUD TRANSLATION

item compared	Traditional CAT	Cloud Translation
Pre translation	Artificial random arrangement of translators, manual preparation terminology database	Automatic preparation of translators, corpus and terminology
Task management	Manually assign tasks to translators, the translation task management are done by Email and telephone	Effective implementation of translation process management online
Content requirements	Ordinary	Good
Cost and quality	Hire professional translators, making the translation costs too high	through continuous learning, the cloud platform will improve the quality of translation, and compress the costs

D. The Key Technologies

Compared with the traditional CAT work mode of the individual as a unit, the biggest characteristic of the cloud translation mode is the users often composed of a big community. It is known to all that different translators'

education background, translation level and translation experience are all different, so that their translation habits and translation psychologies are different too. And translation cognitive research shows that different translators have different mental models facing the same text [2]. In other words, the cloud translation platform should take users individual requirements into consideration when aiding translation generation. Wang Jian-de proposed Interactive Multi Strategy Machine Translation System (Abbreviation IHSMTS), that the interaction between translators and machine system is client-server mode. The translation knowledge that the IHSMTS got will be divided into public knowledge and individual knowledge [3]. When aiding translation generation, the IHSMTS will consider such factors as user's personal information stored in the client, translation experience, hours of work, workload parameters and so on. For the IHSMTS has a comprehensive user priority, to a certain extent, the system can achieve the user's state of better understanding and coordination control. Some researchers proposes a computer assisted translation approach based on user's behavior model, in which user's explicit behaviors in the post-editing process are recorded and user's translation decisions are discovered. In this way, the system can dynamically acquire and share user's translation knowledge to improve the quality of aided translation. Experimental results show that the user behavior model built on the post-editing of the first 30% text in a document improves the BLEU score of the translation candidates for the remaining 70% text by 4.9%. The precision of the translation knowledge in user model achieves 94.1% [4]. But Ye Na's experiment of user behavior model only took selected word knowledge into consideration, so the research is one-sided. Based on Ye Na's algorithm, HAN Ya-dong proposed the template for users according to the relative entropy of the translation knowledge base and the user knowledge base. The experiments based on template matching show that the user model greatly improves the efficiency of collaborative translation [5], [6].

For the aiding translation generation may not match the translation quality, translators should do post-editing, that to modify the error in the aiding translation generation to ensure that the target translation achieves a certain quality requirement [6]. In the process of post-editing the translators generated a large number of behavioral data, which is very valuable clues that can reflect the human thinking process in translation activities and translation difficulties. So the users' post-editing behavior can be used as a feedback to optimize the translation engine, to automatically modify some of the repeatability error or predictable error, and to improve the quality of the auxiliary translation [7]. Traditional CAT system provide translators with fixed post-editing environment and strategy, but the cloud translation platform should provide collaborative community with the post-editing environment, which have different requirements for the auxiliary tools, the assistant way and the interface style. And the platform should be able to act according to the user the interactive behavior and the translation condition, forecast the different users in the different translation stages to the post-editing assistance pattern different demand, the systematic assistance by the

appropriate way, and provide in the appropriate opportunity to the user.

One of the features and advantages of the traditional CAT software is that the system and users can interact with each other, the users can provide the system's auxiliary translation generation with verification, proofreading, etc. The system should make full use of this kind of feedback information, improve the translation knowledge system, optimize the performance of translation, and avoid repeating the same error. Learning from feedback is one of the important factors to reflect the performance of the CAT system. The cloud translation system also can learn translation knowledge from the users' feedback, and the system can learn more translation knowledge from the users' community, and the accumulation of knowledge is more efficient than traditional single user. However this feature can also cause the problem of conflict between different users' translation knowledge. So the cloud system should also have the ability to judge the credibility of translation knowledge, only accept the most credible translation knowledge, and improve the accuracy of translation knowledge [8].

Based on the conclusion and reflection of the tough development of Machine Translation, Zhang Gui-dong presents a new thought of the integration of Knowledge Management and Machine Translation, which centralizes on a user model. The user model is composed of user behavior model and user state model. The user behavior model describes the action and decision of the user's translation process, and improves the efficiency of the accumulation and application of translation knowledge. The state model describes the characteristics and attributes of the user, which is effective for task allocation, collaborative quality control and translation knowledge. The platform based on user model and state model applied to the translation of large scale scientific and technological information and obtained remarkable application effect [8].

III. CORRELATION RESEARCH ACCOMPLISHMENT

With the rapid development of translation industry, organizing multiple spatially distributed translators to complete one large translation task has become a common phenomenon. How to transform single-user, small-scale translation working-mode into multiple-users and large-scale mode is the problem many researchers thought. With the popularization of cloud computing technology, the combination of traditional translation techniques and cloud computing technology has been an important change direction of the translation technology mode in recent years. Under this trend, the "12th Five Year" National Science and Technology Planning Project Collection Guide also launched "coordinated multilingual translation service based on cloud platform and its application" and other related issues. Many universities' research institutions and translation enterprises such as Shenyang Aerospace University, Xiamen University, Huazhong University of Science and Technology, the Trans Network Technology Co., Ltd have engaged in the study of cloud translation platform technology in recent years. Some initial results have been made in the research of the auxiliary translation

generation based on the cloud computing platform:

In July 2008, Beijing, Zhang's research has been identified by the Identification Committee organized by Chinese Information Processing Society, and the committee agreed and announced that "GE-Soft" has successfully accomplished their work: the Cooperative Translation Platform based on Knowledge Management and Intelligent Control Technology, which is developed from their national 863 project of the Integration of Machine Translation and Knowledge Management. The establishment and application of the user model can make the efficiency of human computer cooperation and the user's ability to translate in the collaborative translation system. But there are not further needs analysis and refinement in the GE-soft, especially in user model and translation psychology and behavior learning [8].

Shi Xiaodong and He Zhonghao introduced the Maximum Entropy Based BTG mode, terms detection and recognition technology. He Zhonghao recognized terms detection as a classification problem, in consideration of the existence of professional terms and the number of unprofessional terms in the document, introduced the combination of ensemble learning and multiple classifier to minimize the classification error. Furthermore, HE introduce an aided translation input method into the project team management of the collaborative translation tools based on the cloud translation platform, to better coordinate the relationship between the different translation projects and translators. But the input method is still limited to the local users, and the system does not solve such factors the users' machine performance, local dictionary size restrictions and so on. So the convenience and efficiency of He Zhonghao's algorithm have a lot of room for improvement [9], [10].

Trans Company puts forward the mode of "language network", which is an idea for "private clouds", that Trans' partners can use language network functions and resources, namely to software as a service mode of platform resources available to all partners. In other words, it is very advantageous for the enterprises, especially small and medium-sized translation enterprises, to reduce the initial investment cost and the difficulty of the different operation platform. The "language network" was online at 2013, which have gotten good performance at license translation and resume translation. The Trans internet company also launched the "huoyun term" to public, a free cloud translation software, that can provide professional interpreters for querying terms, terms sharing, terms noting etc. The current "huoyun term" cloud terminology database term has reached 14 million entries, covering more than 200 industry and support the mutual conversion between 56 languages, is the largest and most complete terminology database in China [11], [12], [14]-[16].

At present, the technology of the electronic dictionary based on cloud platform has gradually matured and has been applied. Microsoft's Bing dictionary and NetEase's youdao dictionary are popular too in China. Those cloud dictionaries are based on cloud architecture and cloud service model, which can synchronize the search for a large number of new words on the Internet. With the help of the cloud dictionaries, users can achieve more accurate translation about the words and sentences online. And it can

be easy for Internet users to browse the English website, and read 80% of the web content. But it cannot be the carrier of high quality translation and the best performing for translation.

After more than ten years of research and development, Loto Online Technology Co., Ltd developed a new type of multilingual information retrieval based on cloud computing technology, namely translation cloud massive languages information retrieval engine. The engine can be well docked with the existing CAT software of any kind. The engine can cooperate with CAT systems' probability algorithm model or statistics model, and can get significant reliable translation results. With the help of the cloud engine, the translation efficiency will reach a new level. The matching efficiency of the CAT software, which has some terminology storage and corpus, can be increased nearly doubled. And Loto Company is a leading provider of people.com.cn, CCTV.com as the central focus of the news media translation service. The Loto company also developed the hardware chip specifically for the research and development of semantic recognition arithmetic, which makes it is possible that every server can take massive data operations per second [12], [13].

IV. CONCLUSION AND FUTURE WORK

Under the globalization background, the need of exchange and fusion between different languages has greatly promoted the development of the translation industry. The cloud translation is composed of information technology, network, corpus, terminology storehouse, machine translation, translation professions, and optical character recognition and so on. The cloud translation has greatly improved the efficiency of translation industry, and made it possible to short time processing large amounts of text. The domestic research of computer aided translation based on cloud computing platform is still in its initial stage. Although the technology of electronic dictionary based on cloud platform has gradually matured, there is no comprehensive and in-depth study on the auxiliary translation generation and systematic feedback learning strategy. And the online translation system based on mobile phone platform, which is combined with mobile cloud computing related technologies, is still in an embryonic stage.

Although the cloud translation system has moved towards large-scale real application, but some core technology still needs to be further studied and explored. It can be predicted that such technologies, which are the translation psychology, translator roles control technology, and the cloud translation based on mobile platform will become the focus of future

research about the cloud translation system.

REFERENCES

- [1] L. Bowker, *Computer-Aided Translation Technology: A Practical Introduction*, University of Ottawa Press, Canada, 2001, pp.11-30.
- [2] Y. Linhai, *Translation of Cognitive Psychology*, Science Press, Beijing, China, 2008.
- [3] J. D. Wang, "Design and implementation of multi user interactive translation system based on cooperative mechanism," in *Proc. 2000 of the Sixth Graduate Seminar in Computer Science and Technology*, 2000, pp.134-140.
- [4] Y. Na, G. P. Zhang, Y. D. Han, and D. F. Cai, "Computer assisted translation based on user behavior model," *Journal of Chinese Information Processing*, vol. 25, pp. 98–103, May 2011.
- [5] Y. D. Han, "Research on aiding translation generation in collaborative Translation environment," *Journal of Shenyang Aerospace University*, vol. 28, pp. 60-63, Oct 2011.
- [6] S. O'Brien, "Methodologies for measuring the correlations between post-editing effort and machine text translatability," *Machine Translation*, vol.19, pp. 37-58, Jan 2005.
- [7] J. Doyon, "Automated machine translation improvement through post-editing techniques: analyst and translator experiments," in *Proc. 2000 of the 8th AMTA*, 2008, pp. 346-35.
- [8] G. P. Zhang and D. F. Cai, "A cooperative translation platform based on knowledge management and intelligent control-the integration of knowledge management and computer translation," *Journal of Chinese Information Processing*, vol. 22, pp. 3-11, Sep 2008.
- [9] Z. H. He, J. S. Su, X. D. Shi, Y. D. Chen, and Y. Z. Huang, "An ensemble learning method for maximum entropy based phrase reordering model," *Journal of Chinese Information Processing*, vol. 28, pp. 87-93, Jan 2014.
- [10] Z. H. He, "Research on collaborative translation tools based on cloud translation platform," M.S. thesis, Dept. Pattern recognition and intelligent system, Xiamen University, Xiamen, China, 2014.
- [11] X. L. Li and P. C. Xiao, "Cloud translation display ultra-strong communication ability," *Hubei Daily*, April 17, 2012.
- [12] C. Yi, "Localized translation in the globalization campaign," *Journal of Northeast Normal University (Philosophy and Social Sciences)*, vol. 4, pp. 157-159, April 2011.
- [13] J. Pei, "Cloud translation engine can improve the efficiency of translation," *Chinese Translation Occupation Exchange Congress*, Peking University Press, 2010.
- [14] L. L. Yan, "Language networks started to create a new model of language service," *Chinese Science and Technology Translation Journal*, vol. 26, no. 3, p. 64, March 2013.
- [15] C. L. Wang, "The translation project management and professional interpreter train," *Chinese Translators Journal*, vol. 32, no. 1, pp. 55-59, Jan 2011.
- [16] L. Le and L. L. Yan, *Translation Project Management*, National Defence Industry Press, Beijing, China 2014, ch. 9.



Ren Xiaoyan was born in Jinmen city, Hubei province, China, in 1976. She graduated from China Three Georges University, where is in Yichang city, Hubei province, China, and received her master's degree in engineering. She works currently as a lecturer in the College of Information Technology and Computer Science, China Three Gorges University. The author's research interests include information processing and computer-aided translation. The author has published more than 20 professional papers.