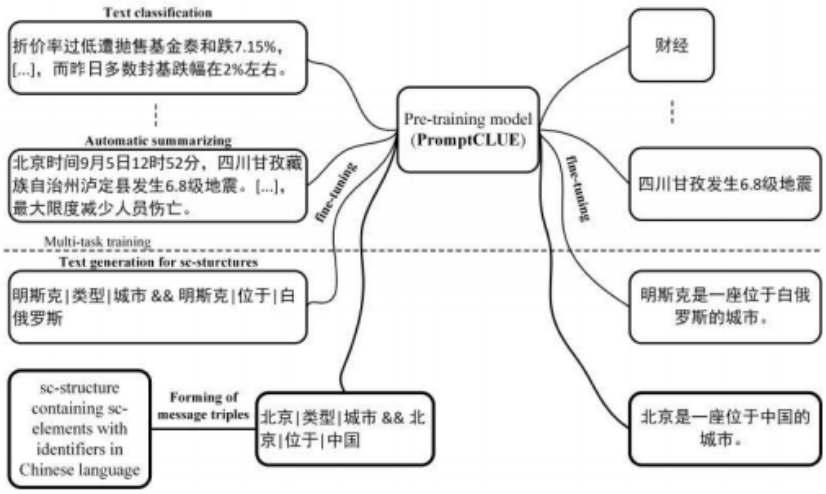

Figure 8: Diagram for text generation based on the pre-training model PromptCLUE

example, when generating the resulted Chinese sentence, the refering expression of each segmentation unit means the final form of the segmentation unit in the resulted generated Chinese sentence. According to the constructed template, the referring expression of the segmentation unit "图 (graph)" is the subject. The segmentation unit "临界图 (critical graph)" is considered as object. The generated Chinese declarative sentence describes "图 (graph) 包含 (inclusion) 临界图 (critical graph)."

Within Technology OSTIS framework some relations are already predefined in the IMS system for the development of ostis-systems, for example "inclusion*", "equivalence*" and so on. For these finite relations (we call domain-independent relations), templates is suitable for text generation. However in various subject domains there are a large amount of infinite relations. In this case, the neural network models can be integrated into the problem solver. In the Figure 8 shown the process of using pre-training model and fine-tuning paradigm to solve the tasks of generation Chinese sentence from sc-structure.

For task of generation Chinese sentence from sc-structure we use the pre-training PromptCLUE, which uses an encoder-decoder architecture using Transformer model and is pre-trained on several sets of Chinese language processing tasks using a huge Chinese corpus (hundreds GB of Chinese corpus) [15]. Afterwards we fine-tuned this model using task of generation Chinese texts from sc-structure (constructed dataset in from of message triple/Chinese sentence pairs). After fine-tuning on the PromptCLUE model, our retrained model can be used to generate Chinese texts from pre-processed sc-structures into message triples.

C. evaluation for Chinese language interface

In order to prove the effectiveness of the sc-model of natural language interface within OSTIS Technology framework, the developed Chinese language interface is currently being evaluated mainly in the following three aspects:

- evaluation of the knowledge base on Chinese language processing;
- evaluation of the efficiency of sc-structures generation;
- evaluation of the Chinese texts generation.

In order to compare the knowledge base on Chinese language processing with other similar existing knowledge bases used for Chinese language processing, the following proposed criteria for comparison of knowledge base within OSTIS Technology framework are highlighted:

- form of knowledge base structuring;
- independence of subject domains from each other;
- form of knowledge representation and form of knowledge storage in the knowledge base;
- possibility to solve problems using logical statements;
- presence of means to visualize the knowledge base.

In Table. I shown the result of comparing the developed knowledge base on Chinese language processing with other knowledge bases about Chinese language processing according the selected criteria.

In principle, on the basis of sc-model of knowledge base in natural language interface, linguistic knowledge at various levels based on existing knowledge base can be integrated in unified knowledge base on Chinese language processing. In addition, the various extraction rules or

Table I: Evaluation of knowledge base on Chinese language processing

Knowledge base	Criteria					
	structured representation and storage of knowledge	subject domain of words	subject domain of sentences	linguistic knowledge on phrases and others	presence and use of logical statements to solve problems	presence of means to visualize the knowledge base
Grammatical KB of Contemporary (GKB)	+	+	-	-	-	-
Mandarin VerbNet	+/-	+	-	-	-	-
HowNet	+	+	+	-	-	-
Chinese Treebank 8.0	-/+	-	+	-	-	-
Knowledge base on Chinese language processing	+	+	+	+	+	+

templates for Chinese texts generation also can be built in knowledge base on Chinese language processing. This advantage is completely absent from other knowledge bases. Moreover, the developed knowledge base on Chinese language processing is structured into the respective subject domains. Sufficient independence between subject domains allows team development, which significantly reduces the time and labor costs in developing a knowledge base compared to developing other knowledge bases.

To evaluate the efficiency of Chinese text analysis (conversion Chinese texts into knowledge base fragments), the ideal way is to calculate the similarity between the sc-structure existing in the knowledge base and the corresponding sc-structure generated by the problem solver of Chinese text analysis. In our situation, the sc-structure is a graphical structure with identifiers in Chinese language. In [16], an approach was proposed for calculating the similarity between semantic graphs (sc-structures), focused on checking the answer to the target question. Therefore the approach can be used to calculate the similarity between the sc-structure existing in the knowledge base and the corresponding sc-structure.

However, this approach does not take into account the influence of the identifiers of each element in sc-structures. The additional metric exact matching is always used for evaluating the effectiveness of knowledge acquisition. The exact matching means that the identifiers of each extracted element (named entities and relations between them) must exactly match the identifiers of the element in knowledge base. To calculate the similarities between the standard sc-structures in knowledge base and the scstructures generated by the problem solver of Chinese text analysis, we manually selected several different kinds of sc-structures.

Table II: Evaluation of similarities between sc-structures

	Three-element construction	Five-element construction	Non-standard construction	Total
Number	15	15	10	40
Average similarity score	0.8125	0.8387	0.7273	**0.7928**

In Table. II shown the results. Depending on the complexity of the sc-structures, the different numbers for different types of sc-structures is selected, then calculate the average similarity score for these sc-structures, finally calculate the overall similarity score to evaluate the efficiency of the problem solver.

As can be seen from Table. II, as the complexity of scstructures increases, the similarity score decreases. Overall the developed problem solver still achieves a relatively good result.

Table III: Evaluation of exact matching of identifiers

	Precision	Recall	F1
Problem solver of Chinese text analysis	0.8289	0.7875	**0.8076**
CORE	0.8308	0.6750	0.7448

According to the metric exact matching, the identifier of each element of selected sc-structures was manually added in Chinese language by trained native Chinese speakers and verified by others. Currently there is the CORE system [17] that basically extracts structure in RDF from Chinese sentences. Therefore for metric exact matching, the CORE system can be used to evaluate the performance of the developed problem solver of Chinese text analysis.

In Table. III shown the experimental results. In summary, the results show that the use of series of Chinese text analysis and constructed extraction rules is effective in extracting knowledge base fragments without any specific human intervention.

To evaluate the generated Chinese texts, in other text generation systems, automatic metrics BLEU-4 [18] and ROUGE-L [19] scores are commonly used to evaluate the quality of generated texts. To evaluate the quality of the generated Chinese texts, the corresponding reference Chinese sentences corresponding to several various types of sc-structures are built manually by trained native Chinese speakers and verified by others.

Currently, there is only Melbourne's best WebNLG system for generating English texts, which is focused on generating English sentences from knowledge base fragments in form of RDF [20]. With the advent of the pre-training model, WebNLG provides a basic system implemented on pre-training model T5 for generating English texts [21]. Without other Chinese text generation systems to compare performance, therefore the performance of developed problem solver of Chinese text generation and other generation systems for English language in the same evaluation metrics BLEU-4 and ROUGE-L are shown in Table. IV and Table. V separately.

Table IV: Evaluation of efficiency for Chinese text generation

	BLEU-4	ROUGE-L
Problem solver of Chinese text generation	**0.5885**	**0.6793**

Table V: Evaluation of efficiency for generation systems for English language

	BLEU-4	ROUGE-L
T5-baseline	0.5520	0.6543
Melbourne	0.5452	0.6350

As can be seen from Table. IV and Table. V, although the generation systems for English language is oriented on generating English language texts from knowledge base fragments in form of RDF, with the help of the combined use of neural network models and semantic models for generating Chinese texts, the developed problem solver achieved relatively promising BLEU-4 and ROUGE-L scores on Chinese texts generation. Moreover experimental results show that the developed problem solver is more suitable for generating Chinese texts when developing interface of ostis-systems.

V. CONCLUSION

This article had proposed a unified semantic model of natural language interface for intelligent system, oriented on conversion natural language texts into knowledge base fragments and generation natural language texts from knowledge base fragments within OSTIS Technology framework. The proposed semantic model of natural language interface mainly consists of sc-model of knowledge base of linguistics, in which the linguistic knowledge at various levels can be constructed, as well as sc-model of problem solvers, which have ability of deeply integrating logical models on rules and neural network models for natural language texts conversion and texts generation using multi-agent approach. Moreover on the basis of the unified semantic model of natural language interface the Chinese language interface of intelligent system in specific subject domains can be implemented with help of developed knowledge base on Chinese language processing and corresponding specific problem solvers for Chinese language processing. In order to verify the performance of the semantic model of natural language interface, we evaluated the developed Chinese language interface in three aspects. According to evaluated results the developed knowledge base on Chinese language processing has ability to integrate various linguistic knowledge for Chinese language processing. Compared to other systems (in these system factual knowledge is represented in form of RDF) for knowledge extraction and text generation, developed corresponding problem solvers could achieve relatively promising scores on specific metrics.

REFERENCES

[1] Y. C. Liu Survey on Domain Knowledge Graph Research. Computer Systems Applications, 2020, vol. 26, No. 06, pp. 1–12.4

[2] L. W. Qian Ontological Approach to the development of natural language interface for intelligent computer systems. *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh system [Open semantic technologies for intelligent systems]*, Minsk, 2022, pp. 217–238.

[3] Q. Liu, Y. Li, H. Duan, Y. Liu, Z. G. Qin Knowledge Graph Construction Techniques. *Journal of Computer Research and Development*, 2016, vol. 53, No. 03, pp. 582–600.

[4] Y. Yang, Z. Wu, Y. Yang, S. Lian, F. Guo, Z. Wang A Survey of Information Extraction Based on Deep Learning. *Journal of Software*, 2019, vol. 30, No. 06, pp. 1793–1818.

[5] Y. Gao Review of Open Domain Information Extraction. *Modern Computer*, 2021, No. 07, pp. 103–110.

[6] A. Fader, S. Soderland, O. Etzioni Identifying Relations for Open Information Extraction. Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, John McIntyre Conference Centre, Edinburgh, 27-31 July 2011, pp. 1535–1545.