# Ontological Approach to Chinese Language Interface Development in Intelligent Systems

Longwei Qian

*Department of Intelligent Information Technology*
*Belarusian State University of Informatics and Radioelectronics*
Minsk, Belarus
Email: qianlw1226@gmail.com

***Abstract*—This article is devoted to development of a unified semantic model of natural language interface, which allows combine various linguistic knowledge on natural language processing into a single knowledge base, as well as deep integration of logical models on rules, neural network models and other problem solving models for natural language processing towards solving conversion natural language texts into knowledge base fragments and generation natural language texts from knowledge base fragments. Moreover main principles of building Chinese language interface is described on the basis of unified model of natural language interface. Finally the developed Chinese language interface is evaluated in order to prove the effectiveness of unified semantic model.**

***Keywords*—ontology, knowledge-based system, knowledge acquisition, text generation, Chinese language processing**

## I. Introduction

With the development of knowledge engineering technology, knowledge-based intelligent systems are actively developed in the direction of solving various complex tasks (for example, natural language processing, images analysis, speech analysis and so on). Moreover the knowledge-based question answering system considered as the next generation search engine is the top research topic in the research area of knowledge-based intelligent systems. As a key component of knowledge-based intelligent systems, the natural language interface is oriented to achieve information exchange between human users and knowledge base of intelligent systems in natural language texts (especially declarative sentences). With the development of intelligent systems, the need to conversion between natural language texts and knowledge base through natural language interface is increasingly prominent. In framework of this article we mainly consider the task of ***conversion input natural language texts into knowledge base fragments*** and ***generation natural language texts from knowledge base fragments*** in natural language interface.

Both conversion natural language texts and generation natural language texts are considered subtasks of natural language processing. The natural language processing is a kind of so-called complex problem, the research of which has been a hot topic in the research area of artificial intelligence.

## II. Related works

In accordance with the types of processed natural language and the range of knowledge base of intelligent systems, in our works natural language interface of intelligent systems is divided into following four classes:

- natural language interface that is independent of the specific natural language and the specific domain;
- natural language interface that is dependent on the particular natural language, but is independent of the particular domain;
- natural language interface that is independent of the particular natural language, but is dependent on the particular domain;
- natural language interface that is dependent on the particular natural language, and as well as is dependent on the particular domains;

Due to the time-consuming and laborious development of world-wide knowledge base and its narrow application scenarios [1], natural language interface that is independent of the particular natural language, but is dependent on the particular domain is the main object of our research in this article. The more detailed description about classes of natural language interface can be seen in [2].

The task of conversion natural language texts into knowledge base fragments is considered as the factual knowledge extraction. In our works this task solved in the natural language interface refers to obtaining factual knowledge (mainly named entities and relations between them) from natural language texts, then formed in the knowledge representation.

In the early stage the factual knowledge extraction is performed in the closed domains, which often requires predefined types of named entities and relations between them. The factual knowledge is extracted from natural language texts using artificial constructed templates and rules, then is formalized in RDF [3]. Nowadays various large language modelings with the help of neural network models are applied for factual knowledge extraction

[4]. However the training of large language models requires expensive hardware equipment and a large of training corpus with annotations, which are time-consuming and laborious to obtain and construct. The factual knowledge extraction using rules doesn't require a huge of training corpus, but the construction of rules manually is inefficient. The factual knowledge extraction from open domains is to extract factual knowledge without requiring a predetermined vocabulary to define the types of named entities and relations between them [5]. This task solution uses a common syntax and lexical constraints for extracting factual knowledge from natural language texts [6]. Early systems are focused on extracting factual knowledge from English sentences. For other natural languages, such as Chinese, the structure and the grammatical features of English language usually are not totally applied for Chinese language. Therefore it is necessary to consider the characteristics of the specific language when extracting factual knowledge. Moreover, when developing factual knowledge extraction systems there is no unified basis, which leads to time-consuming and laborious for systems development.

For generation natural language texts from knowledge base fragments, in the early text generation systems widely use templates and grammar rules to generate natural language texts from structured data (for example, tabular data, fragments in RDF, fragments in OWL and others) [7], [8]. However, the constructed rules and templates are highly dependent on the specific application fields and specific natural language. The construction of rules and templates requires significant amount of labor and time. Recently the large language modelings also are applied for text generation in various applications (for example, text summary, caption generation, text generation from RDF and others) [9], [10]. In these applications the obtaining and construction of high-quality aligned corpus is a huge challenge in solving text generation tasks using large language models.

For development of natural language interface, in modern knowledge-based intelligent systems, the following problems still need to be considered:

- In modern intelligent systems, the lack of unification for development of natural language interfaces leads to significant overhead costs for integration of various components (e.g. knowledge base on natural language processing, component for factual knowledge acquisition, component for natural language texts generation) in the process of developing natural language interfaces;
- Due to the diversity of natural languages, when solving the factual knowledge extraction from open domains and text generation, it's necessary to take into account the characteristics of specific natural languages. However, the lack of unified model to integrate linguistic knowledge at various levels (e.g. lexical aspect, syntactic aspect and others) for natural language processing into unified knowledge base significantly complicates the construction of various systems with their use;
- Whether conversion natural language texts into knowledge base fragments or generation natural language texts from knowledge base fragments, in modern systems, the solution of two tasks usually requires the use of various types of problem solving models. Existing ontology-based approaches are only applicable to the factual knowledge extraction from closed domains. Moreover the lack of unified principles for using various problem solving models for natural language processing (for example, logical models on rules, neural network models and so on). In turn, when development of specific natural language interface the overhead costs increase.

The solution of two above-mentioned tasks, in generally, requires a combination of various linguistic knowledge on natural language processing and the use of various problem solving models for natural language processing. In this article we proposed to use ontological approach to develop a unified semantic model of natural language interfaces, which has ability to implement factual knowledge extraction and text generation. From the perspective of the model structure, the unified semantic model mainly consists of semantic model of knowledge base of linguistic and semantic model of corresponding problem solver for natural language processing, which effectively combines linguistic knowledge at various levels and integrates various problem solving models for two above-mentioned tasks solution.

## III. PROPOSED APPROACH

The ontological approach within OSTIS Technology framework [11] is proposed to develop the unified semantic model of natural language interface of intelligent systems for solution of conversion natural language texts into knowledge base fragments and generation natural language texts from knowledge base fragments. The intelligent systems developed using the OSTIS technology is called knowledge-driven computer systems (ostis-systems). The SC-code is used as an internal formal language for the semantic representation of knowledge in the memory of ostis-systems and provides a unified version of information encoding and a formal basis for developing model of ostis-systems. Several universal variants of visualization of SC-code [11], such as SCg-code, SCn-code, SCs-code will be shown below. Ontological model of any entity described by SC-code will call sc-model.

Within OSTIS Technology framework, natural language interfaces of the ostis-systems is considered as the specialized ostis-system focusing on solving the specific tasks in natural language interfaces [12] (in our work,

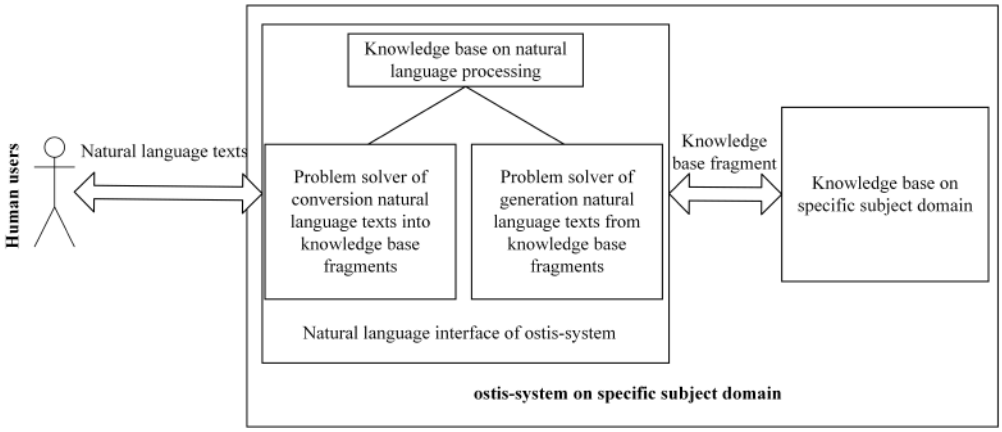


Figure 1: Model of process of natural language processing in the natural language interface

conversion natural language texts into knowledge base fragments and generation natural language texts from knowledge base fragments). The development of sc-model of natural language interface generally includes sc-model of knowledge base (mainly consists of sc-model of knowledge base of linguistic, actions for texts analysis and actions for texts generation) and sc-model of problem solvers of natural language interface. Due to the use of component approach, the development of the entire natural language interface comes down to development and improvement of separate specified components (e.g. knowledge base on natural language processing, component for natural language texts analysis, component for natural language texts generation). The model of the process of natural language processing (Figure 1) in the natural language interface describes the overall process of processing of natural language texts fragments with knowledge base fragments of intelligent systems.

In order to implement conversion natural language texts into knowledge base fragments and generation natural language texts from knowledge base fragments, it is necessary to describe the various linguistic knowledge for natural language processing, the construction of extraction rules, rules and templates for text generation. The development of SC-model of knowledge base of linguistic is to consider structure of knowledge base as a hierarchical system of subject domains and their corresponding ontologies, shown below in SCn-code:

***SC-model of knowledge base of linguistics***

:= [SC-model of knowledge base on natural language processing]

⇐ *section decomposition**:

{• *Subject domain of lexical analysis*

• *Subject domain of syntactic analysis*

• *Subject domain of semantic analysis*

}

These subject domains describe specification of linguistic knowledge at various levels (for example, knowledge on lexical analysis, syntactic analysis, as well as extraction rules, templates for text generation and others) respectively. Usually the knowledge base on natural language processing is not built from scratch. In this work, SC-model of knowledge base is built taking into account existing knowledge bases, such as Wordnet, Verbnet, Treebank, Mandarin VerbNet, Chinese Treebank and others.

In addition to the various type of linguistic knowledge provided in the knowledge base of linguistic, the natural language interfaces should perform some *actions* to solve corresponding tasks. Within OSTIS Technology framework each *internal action of ostis-systems* denotes some transformation performed by some sc-agent (or a group of sc-agents). Therefore when discussing SC-model of actions, we can consider corresponding sc-model of problem solvers. Within OSTIS Technology framework, sc-model of problem solvers is developed as a hierarchical system of agents (sc-agents) [13]. Such approach provides the flexibility and modularity of developed sc-agents, as well as provides the ability to integrate various problem solving models corresponding to these developed sc-agents. In term of abstract sc-agent, the abstract sc-agent is a certain class of functionally equivalent sc-agents, various items of which can be implemented in different ways to specific problems in different programming languages [13].

***SC-model of problem solvers of natural language interface***

⇐ *decomposition**:

{• *SC-model of problem solver for conversion natural language texts into knowledge base fragments*

• *SC-model of problem solver for generation natural language texts from knowledge base fragments*

}

Let us consider the main structure of SC-model of problem solver for conversion natural language texts into knowledge base fragments and generation natural language texts from knowledge base fragments in natural language interfaces of ostis-systems in SCn-code, respectively:

***SC-model of problem solver for conversion natural language texts into knowledge base fragments***

:= [SC-model of problem solver for natural language texts analysis]

⇐ *decomposition of an abstract sc-agent**:

{• *Abstract sc-agent of lexical analysis*

⇐ *decomposition of an abstract sc-agent**:

{• *Abstract sc-agent of decomposing texts into segmentation units*

• *Abstract sc-agent of marking up segmentation units*

}

• *Abstract sc-agent of syntactic analysis*

• *Abstract sc-agent of semantic analysis*

• *Abstract sc-agent of extracting factual knowledge structures into the knowledge base*

• *Abstract sc-agent of logical inference*

}

The SC-model of problem solver for natural language texts analysis is constructed on the basis of the proposed following process for factual knowledge acquisition:

- natural language text is loaded into the interface;
- lexical analysis and syntactic analysis of the input natural language text is performed;
- named entities and relations between them is extracted based on the analyzed syntactic structure and extraction rules.

In principle, this SC-model of problem solver can potentially extract structured knowledge (generally, named entities and relations between them) from texts in different language into the knowledge base of the ostis-systems for a specific subject domain, but the construction of knowledge base on the specific natural language processing, which includes rules for specific natural language processing and extraction rules, will become more complex. In turn overhead costs of construction will increase.

For generation natural language texts from knowledge base fragments the classical pipeline of natural language generation is used as the basis to develop the SC-model of problem solver for generation natural language texts from knowledge base fragments. The developed SC-model of problem solver has higher flexibility. For specific natural language, the developed problem solver can be easily modified accordingly.

***SC-model of problem solver for generation natural language texts from knowledge base fragments***

:= [SC-model of problem solver for natural language texts generation]

⇐ *decomposition of an abstract sc-agent**:

{• *Abstract sc-agent determining sc-structure*

• *Abstract sc-agent dividing determined sc-structure into basic sc-constructions*

• *Abstract sc-agent determining the candidate sc-constructions*

• *Abstract sc-agent transferring candidate sc-constructions into message triples*

• *Abstract sc-agent text planning*

• *Abstract sc-agent for micro-planning*

• *Abstract sc-agent for surface realization*

}

The SC-model of problem solver for natural language texts generation is constructed on the basis of the proposed following process for texts generation:

- a specific sc-structure (fragment of knowledge base) is selected in the knowledge base;
- the candidate basic sc-constructions from the sc-structure is determined, then is translated into a message triple (in the form of subject-relation-object);
- the resulted natural language texts is generated from the message triple as output.

It is worth noting that the composition of sc-constructs has sc-arcs that have specific meanings. Therefore sc-constructions with sc-arcs need to be converted into the corresponding message triples in form of text, which is easier to represent in the form of natural language texts.

The developed unified semantic model of natural language interface ensures the flexibility of developing a specific natural language interface and integration of various components (knowledge base on natural language processing, component for conversion natural language texts into knowledge base fragment and component for text generation) in the interface. The development of natural language interface consists in the development of individual components independently of each other. It is flexible to adjust and make extensions of linguistic knowledge and sc-agents for tasks solution in specific natural language interface. The more detailed description about function of each abstract sc-agent can be seen in [2].

## IV. Implementation of Chinese Language interface

On the basis of unified semantic model of natural language interfaces of ostis-systems, we can implement a specific natural language interface of intelligent help systems for various subject domains. In this section we will describe the implementation of the prototype of Chinese language interface of a intelligent help system

about discrete mathematics. For developing Chinese language interface it's necessary to construct knowledge base on Chinese language processing and corresponding problem solvers for conversion Chinese language texts into sc-structures and generation Chinese language texts from sc-structures, which has ability to integrate logical models on rules and neural network models for Chinese language processing. The detailed processing stage of conversion Chinese language texts into sc-structures and generation Chinese language texts from sc-structures will be shown in followings.

### A. *factual knowledge extraction from Chinese language texts*

Currently there are some restrictions for extracting factual knowledge from Chinese language texts:

- the processed Chinese language texts are Chinese declarative sentences;
- there are specific factual knowledge (named entities and relations between them) in the Chinese declarative sentences;
- due to features of Chinese language, the result of decomposition of Chinese declarative sentences into segmentation units greatly influences the factual knowledge extraction.

In this section the general processing stage of conversion Chinese declarative sentence into sc-structure will be shown in the followings.

*Step 1:* From the point of view of OSTIS technology, any natural language text is a file (sc-node with content or so-called sc-file). The Chinese declarative sentence shown in our example is represented in such a node in Fig 2 and describes: "有限集合(the finite set) ， (comma) 严格地(strictly) 包含 (includes) 二元组(pairs)。(full stop)".


Figure 2: The representation of the Chinese sentence

As shown in Fig 2, according to the written tradition of Chinese language texts, Chinese characters are written one after the other and there are no natural gaps between them. As we know, the lexeme is a term commonly used for lexical analysis in European languages processing. However in Chinese language processing the segmentation unit is considered as the smallest unit . In the "Modern Chinese word segmentation standard used for information processing", a word in Chinese language is represented as a segmentation unit. The precise definition of segmentation units is "a basic unit for Chinese language processing with certain semantic or grammatical functions".

*Step 2:* The Chinese declarative sentence is decomposed into separate segmentation units, lexical analysis is carried out. Afterwards syntactic structure or semantic structure of sentence is analysed, the relations between input sentence and divided segment units, as well as between these segment units in sentence are revealed. The analyzed results of input Chinese sentence is shown in the Figure 3.

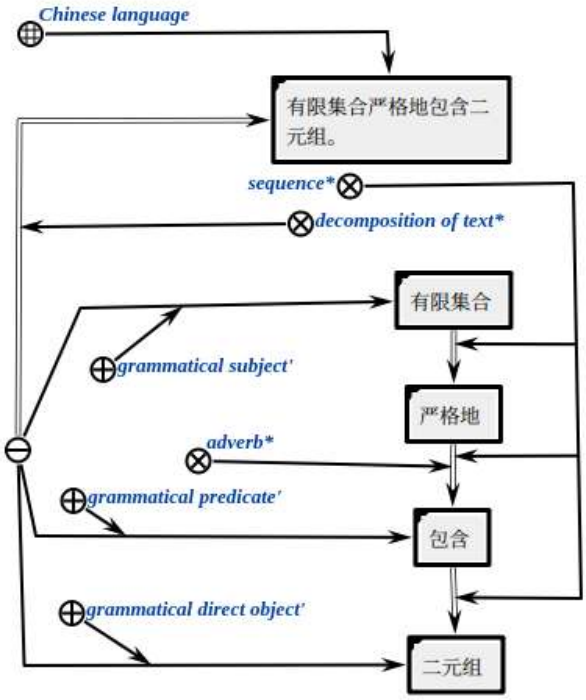


Figure 3: The syntactic structure of input Chinese sentence

*Step 3:* The factual knowledge that mainly consists of named entities and relations between them is extracted based on previous text analysis and extraction rules without contradiction detection. The resulted constructed knowledge base fragment (sc-structure) from input Chinese declarative sentence is shown in the Figure 4.

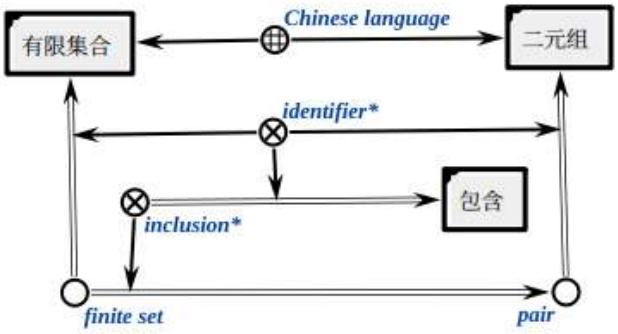


Figure 4: The constructed sc-structure from input Chinese sentence

It is important to note that in this case, a knowledge base fragment can be directly converted into knowledge base without linking extracted named entities and relations between them from the input Chinese sentence with the corresponding entities and relations defined in the knowledge base of intelligent help system.

*B. text generation from knowledge base*

In this section the processing stage of generation Chinese declarative sentence from sc-structure is described. The processing stage is roughly divided into two steps: firstly converting sc-structure from knowledge base into message triples; then generating Chinese declarative sentence from translated message triples. The description about concept message triple can be found in [14].

In our works there are some restrictions for Chinese language texts generation from knowledge base fragments:

- the knowledge base fragment is completed and has sc-elements with identifiers in Chinese language;
- the generated Chinese language texts are Chinese declarative sentences.

*Step 1:* The selected sc-structure is divided into standard basic sc-constructions, afterwards from which the candidate sc-construction is selected and will be converted into "message triple", then into resulted Chinese sentences. A candidate sc-construction (belong to standard basic sc-construction) is shown in SCg (Figure 5). The candidate sc-construction contains sc-elements with identifiers in Chinese language. Identifiers in Chinese language of each sc-element of sc-construction have corresponding specific segmentation units of Chinese language.

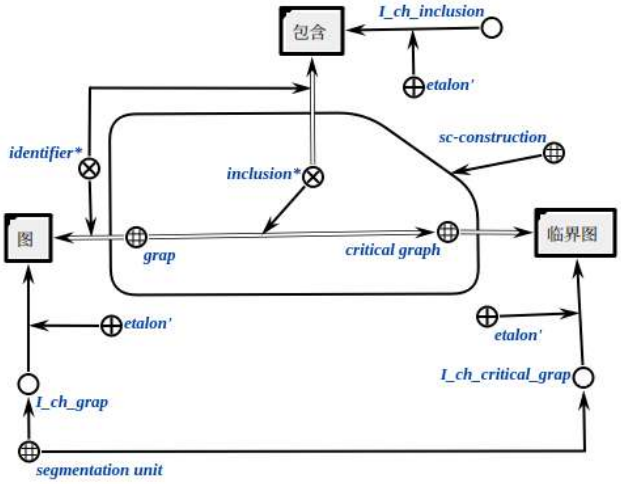


Figure 5: The determination of candidate sc-construction

*Step 2:* The candidate sc-construction is transferred to message triple. The converted message triple consists of sc-files (sc-node with content) containing segmentation units written by trained native Chinese speakers and verified by others. The message triple that corresponds to candidate sc-construction is generated in the Figure 6, in which each sc-element is a file corresponding to a certain segment unit in Chinese language. The contents of some sc-files (e.g. "临界图(critical graph)" ) correspond to the identifier of sc-element in the sc-construction, meanwhile the contents of some sc-files are added when building message triple.

It is important to note that relation of each message triple is the core. Sometimes the relation represents the specific meaning of sc-arc or sc-edge in the sc-structure in form of texts. The main task of text generation is to find suitable text fragment to explain the relation of each message triple in order to generate fluent texts. In general, the subject and object of each message triple are kept constant.

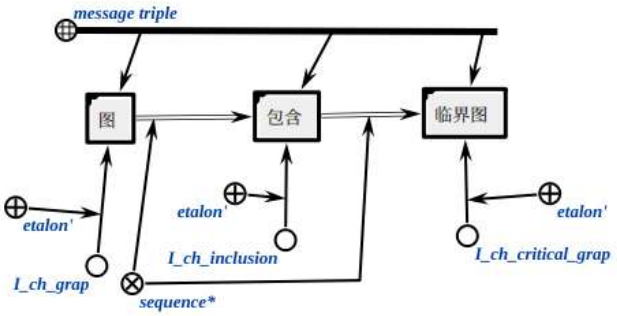


Figure 6: The message triple for candidate sc-construction

*Step 3:* Finally the sc-files are concatenated with certain form of that segment units to generate the resulting Chinese narrative sentence according to the permissible sequence on the constructed template for message triple with the relation "inclusion" (Figure 7). When generating result texts for some natural languages, word forms are changed according to syntactic rules (e.g. capitalizing the first word in a sentence, subject–verb agreement and others), and then added to the result texts. The relation *reference expression** is a quasi-binary relation, connecting a word to its combinatory variants.

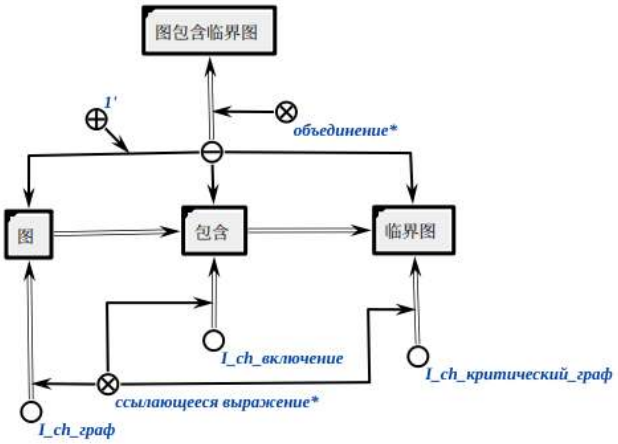


Figure 7: The generated Chinese declarative sentence

For some European languages, The inflected form of the lexical units in sc-files (e.g. singular or plural and other inflected forms) is expressed in the resulted generated texts according to the syntactic rules of a particular natural language. However, due to features of Chinese language, the processing of this step is relatively easier. In this

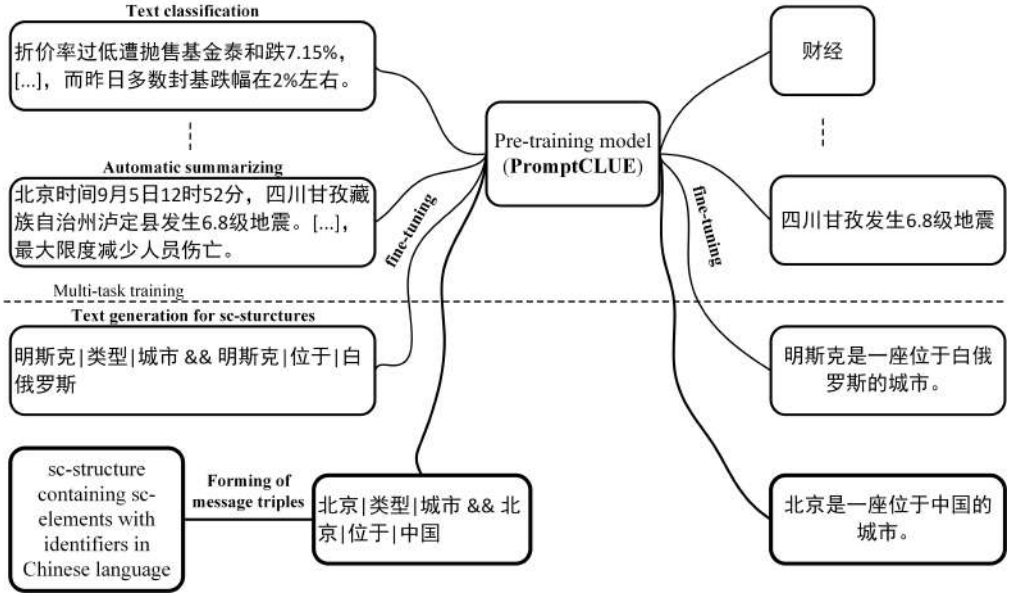


Figure 8: Diagram for text generation based on the pre-training model PromptCLUE

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 Verb-Net | +/- | + | - | - | - | - |
| 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 sc-structures 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 sc-structures 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.

[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.

[7] I. Androutsopoulos, G. Lampouras, D. Galanis Generating Natural Language Descriptions from OWL Ontologies: the NaturalOWL System. *Journal of Artificial Intelligence Research*, 2013, vol. 48, No. 01, pp. 671–715.
[8] A. Gatt, E. Krahmer Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research*, 2018, vol. 61, pp. 65–170.
[9] K. Hu, X. F. Xi, Z. M. Cui, Y. Y. Zhou, Y. J. Qiu Survey of Deep Learning Table-to-Text Generation. *Journal of Frontiers of Computer Science and Technology*, 2022, vol. 16, No. 11, pp. 2487–2504.
[10] G. Claire, S. Anastasia, N. Shashi The WebNLG Challenge: Generating Text from RDF Data. In Proceedings of the 10th International Conference on Natural Language Generation, Santiago de Compostela, Spain, 2017, pp. 124–133.
[11] V. V. Golenkov, N. A. Gulyakina Proekt otkrytoi semanticheskoi tekhnologii komponentnogo proektirovaniya intellektual'nykh sistem. Chast' 1 Printsipy sozdaniya [Project of open semantic technology of component designing of intelligent systems. Part 1 Principles of creation]. *Ontologiya proektirovaniya [Ontology of designing]*, 2014, No. 1, pp. 42–64 (In Russ.).
[12] M. E. Sadouski The structure of next-generation intelligent computer system interfaces. *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh system [Open semantic technologies for intelligent systems]*, Minsk, 2022, pp. 199–208.
[13] D. V. Shunkevich Hybrid problem solvers of intelligent computer systems of a new generation. *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh system [Open semantic technologies for intelligent systems]*, Minsk, 2022, pp. 119–144.
[14] L. W. Qian, W. Z. Li Ontological Approach for Generating Natural Language Texts from Knowledge Base. *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh system [Open semantic technologies for intelligent systems]*, Minsk, 2021, pp. 159–168.
[15] J. Li, H. Hu, X. Zhang, M. Li, L. Li, L. Xu Light Pre-Trained Chinese Language Model for NLP Tasks. In CCF International Conference on Natural Language Processing and Chinese Computing, Minsk, 14 October 2020, Springer, Cham. zhengzhou, 2020, pp. 567–578.
[16] W. Z. Li, L. W. Qian Development of a problem solver for automatic answer verification in the intelligent tutoring systems. *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh system [Open semantic technologies for intelligent systems]*, Minsk, 2021, pp. 169–178.
[17] Y. H. Tseng, L. H. Lee Chinese Open Relation Extraction for Knowledge Acquisition. Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, Gothenburg, Sweden, 26-30 April 2014, pp. 12–16.
[18] K. Papineni, S. Roukos, T. Ward, W. J. Zhu BLEU: a method for automatic evaluation of machine translation. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL '02), USA, 2002, pp. 311–318.
[19] C. Y. Lin ROUGE: A Package for Automatic Evaluation of Summaries. In Text Summarization Branches Out, Barcelona Spain, 2004, pp. 74–81.
[20] B. D. Trisedya, J. Z. Qi, R. Zhang, W. Wang GTR-LSTM: A Triple Encoder for Sentence Generation from RDF Data. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne Australia, 2018, pp. 1627–1637.
[21] M. Kale, A. Rastogi Text-to-Text Pre-Training for Data-to-Text Tasks. In Proceedings of the 13th International Conference on Natural Language Generation, Dublin Ireland, 2020, pp. 97–102.

## Онтологический подход к разработке китайско-языкового интерфейса в интеллектуальных системах

Цянь Лунвэй

В статье рассматриваются существующие подходы к приобретению фактографических знаний из текстов естественного языка и генерации текстов естестфенного языка из фрагментов базы знаний (фактографических знаний), которые рассматриваюся как две основные задачи, решаемые естественно-языковыми интерфейсами интеллектуальнных систем в нашей работе. Был проведен анализ проблем, возникающих при разработке естественно-языкового интерфейса интеллектуальнных систем, а также приобретении фактографических знаний из текстов естественного языка и генерации текстов естественного языка из фрагментов базы знаний в настоящее время.

В рамках технологии OSTIS был предложена разработка единой семантической модели естественно-языкового интерфейса интеллектуальнных систем, которые в основном состоят из sc-модели базы знаний лингвистики и sc-модели соответствующих решателей задач для обработки естественного языка. Среди них в sc-модели базы знаний лингвистики позволяется объединение лингвистических знаний на различных уровнях, в sc-модели соответствующих решателей задач позволяется интеграция моделей на основе правил и моделей нейронных сетей для обработки естественного языка. Более того, на основе единой семантической модели естественно-языкового интерфейса был реальзаван китайско-языковой интерфейс ostis-систем и оценен разработанный китайско-языковой интерфейс по трём аспектам. По сравнению с другими системами, разработанный китайско-языковой интерфейс имеет лучшую эффективность.

Received 29.03.2023