

# Entity Attribute Framework for the Compiling Model of Natural Language Understanding

Honglin Wu<sup>1,a,\*</sup>, Ruoyi Zhou<sup>2,b</sup> and Ke Wang<sup>3,c</sup>

<sup>1</sup>College of Computer Science and Engineering, Northeastern University, Shenyang, China

<sup>2</sup>School of Information Engineering, Zhengzhou University, Zhengzhou, China

<sup>3</sup>Research Center for Artificial Intelligence, Shenyang Linge Technology Co., Ltd., Shenyang, China

<sup>a</sup>wuhl@mail.neu.edu.cn, <sup>b</sup>zhouuo.yi@qq.com, <sup>c</sup>flyingegg.ke@gmail.com

\*Corresponding author

**Keywords:** Compiling model, Natural language understanding, Entity attribute framework.

**Abstract.** Semantic knowledge can be separated from language by the knowledge abstract patterns which use entity attribute relationships. With the existing grammar knowledge, the computer can deal with grammar and semantic analysis step by step. Thus the human understanding process is simulated by the computer. This syntax and semantic separation of processing modes is consistent with the compilation of the computer programming language. In this way, we can accurately define the task of understanding natural language by a model of compiling natural language. What we are concerned with is the framework of knowledge, so that computers can use those knowledge accurately. The framework of knowledge include the definition and the organization of the entities, attributes and relationships of the real world. This paper proposed an entity attribute framework for the compiling model of natural language understanding by computer. Language is the carrier of the real world. The entities and attributes in the real world are simple and static objects, which can be directly mapped to the words. We defined the expression system using language units as entity-attribute-framework expression system. In this expression system, we divide the language units into entity words, attribute words, attribute value words and collection frameworks. In order to discuss that the semantic knowledge would have compiled calculation mode, we also compare these knowledge units with the computer programming language in this paper.

## 1. Introduction

Semantic knowledge can be separated from language by the knowledge abstract patterns which use entity attribute relationships. With the existing grammar knowledge, the computer can deal with grammar and semantic analysis step by step. Thus the human understanding process is simulated by the computer. This syntax and semantic separation of processing modes is consistent with the compilation of the computer programming language. In this way, we can accurately define the task of understanding natural language by a model of compiling natural language.

Now what we are concerned with is the framework of knowledge, so that computers can use those knowledge accurately. The framework of knowledge include the definition and the organization of the entities, attributes and relationships of the real world. This paper proposed an entity attribute framework for the compiling model of natural language understanding by computer.

## 2. The Entity Attribute Framework

Language is the carrier of the real world. The entities and attributes in the real world are simple and static objects, which can be directly mapped to the words. However, the operations and relationships represent the actions of different entities, or entities and their attributes. One action is a complex situation. Unconsidering the internal definition of the situation, the core verb as the identity of the situation would be used to connect the entities or attributes which have relationships.

By the language collocation mode defining the situation verb, the entities and attributes corresponded to the words at the language level would be organized to the network of concept. We defined the expression system using language units as entity-attribute-framework expression system. In this expression system, we divide the language units into entity words, attribute words, attribute value words and collection frameworks. In order to discuss that the semantic knowledge would have compiled calculation mode, we also compare these knowledge units with the computer programming language.

### 3. Entity words

Entity words are the language images of entities. The entity words would include objects that exist objectively or imagine, and also include abstract definitions, events time, places, and so on. One entity word could be a conceptual entity of a category, which corresponding to the top class in the programming language; It also could be an instantiation entity of one class. That entity must inherit from a conceptual entity. It corresponds to the object in the programming language.

### 4. Attribute words and attribute value words

Attribute words are the language images of attributes. One attribute word represents the structure or character of an entity. Meanwhile, one attribute word could be an entity word. Such as that **head** is a part of **people**, which is the attribute of **people**. At the same time, **head** is an entity word, and has its own attributes. One attribute word could not be an entity. Such as **age** of people would not be real object. The attribute words correspond to the member variables defined in the class in the programming language. The data types could be basic data types or another class (entity).

Attribute value words are the language images of the values of attributes. Such as the value of **age** would be a specific number. For an attribute word which also is an entity word, its value would be another instantiation of the entity. Such as **wheel** is an attribute of **car**. The value would be an entity which has its own attributes. Originally, the value of a simple attribute would be a given quantity or a representation of degree concepts.

For instance, **long**, **short** and **1 meter** show the value of **length** attribute of an entity. **1 meter** is a specific number. Then **long** and **short** are representations of degree concepts. The value of abstract attributes is only the representation of degree concept. For example, **fine** and **preciseness** are the values of the abstract attribute of **style**. Different attributes may have the same attribute value intersection. For example, **1 meter** would be not only the value of length, but also the value of the height.

The attribute value words correspond to the value of the member variables in the programming language. For the class type member variable (entity attribute), its value would be a new instantiation of objects. Meanwhile, for the basic data type member variable (simple attribute), its value would be a constant number or a variable constrained by the basic data type as the tokens that could be automatically identify in compiler.

### 5. Collocation Frameworks

Collocation frameworks represent the language collocation model of relationship between entities and entities or attributes. Because the function is to identify the entity or attribute which participates in an action relationship, this collocation framework is a simple language instruction. The core of the frameworks is a verb. A collocation framework uses a collocation template of a verb to represent relationship or action between two entities, two attributes and an entity and an attribute. The verb is a fixed part of the template. The entity words and attribute words are the value of the generalized part of the template.

In that case, this framework could be inherited by the vertical relationship between entities. For instance, the template of “people eat something” is [#entity1] eat/v [#entity2]. Since the entities are generalized, the variable of this framework could be changed by the same type of entities which has the same root in the inherited network.

Thereby, the collocation framework could be inherited used. Corresponding to the programming language, the verb expressed the action is a member function or a friend function. These functions are overloaded with the default parameters. The collocation framework is the declaration of this verb function. Its use is to indicate a legal non-default parameter, such as entities or attributes. The definition of the verb function is the complex descriptive definition of the “situation” mentioned above.

## 6. The Compiling Model of Semantic Knowledge

It would be possible to compile natural language as the specific mapping of the abstract definition and the correspondence between the expression form and the programming language. Not all words can correspond to entities, attributes, or verbs of frameworks. Many words can not find their categories because they are keywords in the natural language compilation system. These words indicate the value of the default parameter in the verbal functions.

Obviously, for the compilation of natural language understanding, the first step is to establish the symbol tables of the sentences or paragraphs. That is mapping the language to the picture. If the definition of *picture* would be complete, the prototype of world would be reproduce. That is the thorough understanding of the language.

Therefore, the basic semantic knowledge constructed for computer should be the knowledge system that would construct the symbol tables automatically. This knowledge system not only has the structure definition of symbol table, but also has the connection point in the process of symbol table establishment. *Situation*, which is the function definition of the verb, would be established based on the natural language symbol table.

## 7. Conclusion

Semantic knowledge can be separated from language by the knowledge abstract patterns which use entity attribute relationships. With the existing grammar knowledge, the computer can deal with grammar and semantic analysis step by step. Thus the human understanding process is simulated by the computer.

This syntax and semantic separation of processing modes is consistent with the compilation of the computer programming language. In this way, we can accurately define the task of understanding natural language by a model of compiling natural language. What we are concerned with is the framework of knowledge, so that computers can use those knowledge accurately.

The framework of knowledge include the definition and the organization of the entities, attributes and relationships of the real world. This paper proposed an entity attribute framework for the compiling model of natural language understanding by computer. Language is the carrier of the real world. The entities and attributes in the real world are simple and static objects, which can be directly mapped to the words.

We defined the expression system using language units as entity-attribute-framework expression system. In this expression system, we divide the language units into entity words, attribute words, attribute value words and collection frameworks. In order to discuss that the semantic knowledge would have compiled calculation mode, we also compare these knowledge units with the computer programming language in this paper.

## Acknowledgement

This research was financially supported by the National Natural Science Foundation of China (61370155). We thank Li Zhang and Siwei Rao for the fundamental research on entity attribute knowledge base construction.

## References

- [1] Y. Yin, GDC: A robust tag recommendation algorithm, *Journal of Computational Information Systems*, vol. 22, pp. 8061-8069, 2015.
- [2] F. Zhang, Z.M. Ma, J.W. Cheng, A survey on fuzzy ontology for the semantic web, *Knowledge Engineering Review*, vol. 3, pp. 1-44, 2016.
- [3] H.L. Wu, R.Y. Zhou, K. Wang, Chinese language analyzing using entity attribute semantic knowledge base, *Proceedings of the 2nd International Symposium on Advances in Electrical, Electronics and Computer Engineering*, vol. 124, pp. 470-473, 2017.
- [4] H.L. Wu, R.Y. Zhou, K. Wang, The structure of domain entity attribute semantic knowledge base for natural language processing, *Proceedings of the International Conference on Computer, Electronics and Communication Engineering*, 2017, unpublished.
- [5] K. Wang, H.L. Wu, Research on neologism detection in entity attribute knowledge acquisition, *Proceedings of the 5th International Conference on Machinery, Materials and Computing Technology*, *Advances in Engineering Research*, vol. 126, pp. 690-693, 2017.
- [6] H.L. Wu, R.Y. Zhou, K. Wang, Extracting attribute words for domain entity knowledge base construction, *Proceedings of the International Conference on Electronics, Electrical Engineering and Information Science*, 2017, unpublished.
- [7] H.L. Wu, R.Y. Zhou, K. Wang, Retrieving collocation frameworks for entity attribute knowledge acquisition, *Proceedings of the 2017 5th International Conference on Frontiers of Manufacturing Science and Measuring Technology*, *Advances in Engineering Research*, vol.130, pp. 1550-1553, 2017.
- [8] H.L. Wu, R.Y. Zhou, K. Wang, Template based attribute value words acquisition in entity attribute knowledge base construction, *Proceedings of the International Conference on Computing Intelligence and Information System*, 2017, unpublished.
- [9] F. Zhang, Z.M. Ma, J.W. Cheng, Enhanced entity-relationship modeling with description logic, *Knowledge-Based Systems*, vol. 93, pp. 12-32, 2015.
- [10] H.L. Wu, Y.Y. Liu, S.M. Liu, A sentence alignment model based on combined clues and kernel extensional matrix matching method, *Proceedings of AASRI Conference on Computational Intelligence an Bioinformatics*, pp. 468-473, 2012.
- [11] H.L. Wu, Y.Y. Liu, and S.M. Liu, Machine translation based on direct matching between text and pattern, *Proceedings of the 2012 International Conference on Computer Science and Service System*, pp. 3751-3754, 2012.
- [12] H.L. Wu and S.M. Liu, Word alignment between Chinese and Japanese using maximum weight matching on bipartite graph, *Proceedings of the 21st International Conference on the Computer Processing of Oriental Languages*, pp. 75-84, 2006.
- [13] H.L. Wu, R.Y. Zhou, K. Wang, Knowledge representation of entity attribute frame for natural language understanding, *Proceedings of the 2nd International Conference on Advances in Management Engineering and Information Technology*, 2017, unpublished.