

# Chinese Language Analyzing Using Entity Attribute Semantic Knowledge Base

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

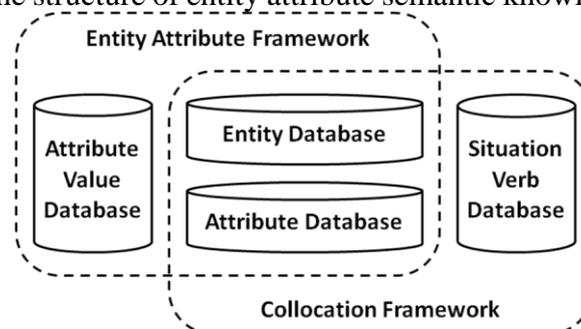
<sup>a</sup>wuhl@mail.neu.edu.cn, <sup>b</sup>zhouuo.yi@qq.com, <sup>c</sup>flyingegg.ke@gmail.com

**Keywords:** Natural Language Analyzing, Entity, Attribute, Knowledge Base.

**Abstract.** A Chinese language analyzing method by using entity attribute semantic knowledge base is proposed in this paper. The knowledge base consists of four parts: the situation verb database, the entity database, the attribute database and the attribute value database. The terms in the above four database are organized into a network according to the relationship between entities and attributes. The analyzing procedure can be generalized into three steps. The first step is to segment the given sentence by matching the terms in the knowledge base; mark the possible semantic categories of every word; specify the possible attribute assignment information for each attribute value word according to the matching relationship between the attribute words and attribute value words. The second step is to identify the entities by the attribute naming rules of instance entities; determine the attribute assignment information of the attribute value words based on the marked attribute information. The third step is to determine the subjects of collocation according to slot information of the situation verb stored in the collocation framework; mark those subjects of collocation. The process of analysis of annotation type is highly consistent with the way of people understanding language. The result of each word of the sentence is consistent with the result of the understanding in the human brain. The proposed method reproduced the complete process of language comprehension, which can be well applied to language analysis.

## Introduction

A Chinese language analyzing method by using entity attribute semantic knowledge base is proposed in this paper. The analyzing procedure can be generalized into three steps: the segmentation step, the identification step and the determination step. The knowledge base by which the sentence is analyzed consisted of four parts: the situation verb database, the entity database, the attribute database and the attribute value database. The structure of entity attribute semantic knowledge base is shown in Fig. 1.



**Fig. 1** The structure of entity attribute semantic knowledge base

The entity attribute framework defined the internal attribute structure, the possible value of each attribute, and the inheritance relationship between entities. A demonstration of the storage definition of entity attribute frame is shown in Fig. 2.

```

<entity>
  <name>DianNao(computer)</name>
  <parent>NULL</parent>
  <attr><name>PinPai(brand)</name><value>SHENZHOU|HP|DELL|.....</value></attr>
  <attr><name>WaiGuan(appearance)</name><value>GaoYa(elegant)|ShiShang(fashion)|.....</value></attr>
  .....
</entity>
<entity>
  <name>^attr[PinPai(brand)]+attr[XiLie(series)]+attr[XingHao(model)]$</name>
  <parent>DianNao(computer)</parent>
</entity>

```

**Fig. 2** Demonstration of the entity attribute framework

The collocation framework defined the collocation structure of verbs with an index of the verb terms. A demonstration of the storage definition of collocation frame is shown in Fig. 3.

```

<operation>
  <verb>CaiYong(use)</verb>
  <frame>
    <template>SLOT_1 CaiYong(use) SLOT_2<template>
    <collocation>(DianNao(computer),SheJi(design))|(DianNao (compter),CPU|.....</collocation>
  </frame>
</operation>
<operation>
  <verb>DaZai(equipped with)</verb>
  <frame>
    <template>SLOT_1 DaZai(equipped with) SLOT_2<template>
    <collocation>(DainNao(computer),CPU)|(DianNao(computer), PingMu(screen)).....</collocation>
  </frame>
</operation>

```

**Fig. 3** Demonstration of the collocation framework

## Sentence Analyzing

The sentence analyzing procedures can be generalized into three steps. Step A: Segment the given sentence by matching the terms in the entity database, the attribute database, the attribute value database and the situation verb database; Mark the possible semantic categories of every word in the given sentence according to the corresponding knowledge databases (entity, attribute, attribute value or situation verb); Specify the possible attribute assignment information for each attribute value word according to the matching relationship between the attribute words and attribute value words.

We use a Chinese sentence as a demonstration of language analyzing procedure. The sentence is “ShenZhou JingDun K480P DaZai BenTeng ShuangHe ChuLiQi JiaGe ShiHui XingNeng QiangDa”. Where “ShenZhou” is a brand of computer; “JingDun” is a series of ShenZhou computer; “DaZai” means “equipped with”; “BenTeng” means “Pentium”; “ShuangHe” means “dual core”; “ChuLiQi” means “CPU”; “JiaGe” means “price”; “ShiHui” means “affordable”; “XingNeng” means “performance”; “QiangDa” means “powerful”. The analyzing result of the given sentence after step A is shown in Fig. 4.

```

ShenZhou/value(PinPai)  JingDun/value(XiLie)  K480P/value(XingHao)
DaZai/verb  BenTeng/value(ChuLiQi)  ShuangHe  ChuLiQi/attr
JiaGe/attr  ShiHui/value(JiaGe|JiaWei)  XingNeng/attr  QiangDa/value(XingNeng)

```

**Fig. 4** Demonstration of sentence analyzing step A

Where “PinPai” means “brand”; “XiLie” means “series”; “XingHao” means “model”; “JiaWei” means “price”. After sentence segmentation, we get three attribute words, six attribute value words and one situation verb and specify the possible attribute assignment information for the six attribute

value word according to the matching relationship between the attribute words and attribute value words.

Step B: Identify the entities by the attribute naming rules of instance entities. Determine the attribute assignment information of the attribute value words based on the marked attribute information. The analyzing result of the given sentence after step B is shown in Fig. 5.

[ShenZhou/value(PinPai) JingDun/value(XiLie) K480P/value(XingHao)]/entity  
 DaZai/verb BenTeng/value(ChuLiQi) ShuangHe ChuLiQi/attr  
 JiaGe/attr ShiHui/value(JiaGe|JiaWei) XingNeng/attr QiangDa/value(XingNeng)

**Fig. 5** Demonstration of sentence analyzing step B

Step C: Determine the subjects of collocation according to slot information of the situation verb stored in the collocation framework; Mark those subjects of collocation. The analyzing result of the given sentence after step C is shown in Fig. 6.

[[[ShenZhou/value(PinPai) JingDun/value(XiLie) K480P/value(XingHao)]/entity]/slot\_1  
 DaZai/verb [BenTeng/value(ChuLiQi) ShuangHe [ChuLiQi/attr]/slot\_2]/relation  
 JiaGe/attr ShiHui/value(JiaGe|JiaWei) XingNeng/attr QiangDa/value(XingNeng)

**Fig. 6** Demonstration of sentence analyzing step C

From the demonstration, we found that the process of analysis of annotation type is highly consistent with the way of people understanding language. The result of each word of the sentence is consistent with the result of the understanding in the human brain. The abstracts of knowledge in this paper reproduced the complete process of language comprehension, which can be well applied to language analysis.

## Conclusion

A Chinese language analyzing method by using entity attribute semantic knowledge base is proposed in this paper. The knowledge base consists of four parts: the situation verb database, the entity database, the attribute database and the attribute value database. The analyzing procedure can be generalized into three steps. The first step is to segment the given sentence by matching the terms in the knowledge base; mark the possible semantic categories of every word; specify the possible attribute assignment information for each attribute value word according to the matching relationship between the attribute words and attribute value words. The second step is to identify the entities by the attribute naming rules of instance entities; determine the attribute assignment information of the attribute value words based on the marked attribute information. The third step is to determine the subjects of collocation according to slot information of the situation verb stored in the collocation framework; mark those subjects of collocation. The process of analysis of annotation type is highly consistent with the way of people understanding language. The result of each word of the sentence is consistent with the result of the understanding in the human brain. The proposed method reproduced the complete process of language comprehension, which can be well applied to language analysis.

## Acknowledgements

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] 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.

- [2] K. Wang, H.L. Wu, Research on neologism detection in entity attribute knowledge acquisition, Proceedings of the 5th International Conference on Computer Science, Electronics Technology and Automation, 2017, unpublished.
- [3] F. Zhang, Z.M. Ma, J.W. Cheng, A survey on fuzzy ontology for the semantic web, Knowledge Engineering Review, 3 (2016) 1-44.
- [4] Y. Yin, GDC: A robust tag recommendation algorithm, Journal of Computational Information Systems, 22 (2015) 8061-8069.
- [5] 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.
- [6] H.L. Wu, R.Y. Zhou, K. Wang, Retrieving collocation frameworks for entity attribute knowledge acquisition, Proceedings of the 3rd International Conference on Electronic Design, Computer Networks and Automated Verification, 2017, unpublished.
- [7] 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.
- [8] F. Zhang, Z.M. Ma, J.W. Cheng, Enhanced entity-relationship modeling with description logic, Knowledge-Based Systems, 93 (2015) 12-32.
- [9] N. Chomsky, G.A. Miller, Introduction to the formal analysis of natural languages, Handbook of Mathematical Psychology, 2 (1962) 269-321.
- [10] C.E Shannon, Communication theory of secrecy systems, Bell System Technical Journal, 28 (1949) 656-715.
- [11] 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, (2012) 468-473.
- [12] 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, (2012) 3751-3754.
- [13] 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, (2006) 75-84.
- [14] S. Abraham, K. Ferenc Kiefer, A Theory of Structural Semantics, Mouton & Co., Hague, 1967.
- [15] 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.