Construction of the Basic Sentence-pattern Instance Database Based on the International Chinese Textbook Treebank

Abstract—In order to better serve the international Chinese teaching, according to the sentence-based syntactic system and Chinese syntax structure characteristics, the paper builds a basic sentence-pattern instance database based on the international Chinese textbook Treebank. Based on the instance database, the paper adds an index for the predicate, and then extracts the relevant knowledge and information of predicate keywords. Finally, a certain amount of complex natural language sentences are converted into more instances with teaching values, which provide basic resources for international Chinese teaching.

Keywords—Sentence-based Syntactic; International Chinese Textbook Treebank; Basic Sentence-pattern Instance; Single-predicate Core

1 Introduction

Jiaxi Li proposed the sentence-based syntactic system in the Chinese grammar, which has a close connection with the teaching grammar system. Based on the sentence-based idea, Language and Language Resources Research Center of Beijing Normal University has constructed the international Chinese textbook Treebank. From the existing international Chinese textbook Treebank the paper extracts sentence-pattern instances with single-predicate core, which belong to the different basic sentence patterns. Although these basic sentence-pattern instances are not complete natural language sentences, they have the teaching value. This paper adds a middle layer between the Treebank and teaching grammar, namely sentence-pattern instance layer. A certain scale Treebank is extended to a larger scale basic sentence-pattern instance database based on the international Chinese textbook Treebank, which provides an idea for the application of the international Chinese textbook Treebank in the international Chinese teaching.

2 Basic Sentence-pattern Instances

The basic sentence-pattern instance database only contains the single-predicate core sentences which reflect the argument structure of the central verb.

Four corresponding basic sentence-pattern instances can be extracted from the (What do you think of the rule that the state does not allow the celebrity to appear in the advertisement as a patient) *

- "What do you think of the rule that the state does not allow the celebrity to appear in the advertisement as a patient"
- "What do you think of the rule that the state does not allow the celebrity to appear in the advertisement as a patient"
- "What do you think of the rule that the state does not allow the celebrity to appear in the advertisement as a patient"
- "What do you think of the rule that the state does not allow the celebrity to appear in the advertisement as a patient"

3 International Chinese Textbook Treebank

4 Extraction Process of the Basic Sentence-pattern Instance

4.1 Extraction Principle of the Basic Sentence-pattern Instance

4.2 Extraction Method of the Basic Sentence-pattern Instance

4.3 Storage of the Basic Sentence-pattern Instance Database

- "S" represents the subject component
- "D" represents adverbial component
- "K" indicates predicate component
- "O" indicates object component

5 Overview of the Basic Sentence-pattern Instance Database

6 Conclusion

Basic sentence-pattern instances are extracted from the international Chinese textbook Treebank. Based on the instance database, through establishing the keyword index, all related knowledge and information about the predicate keywords are extracted. The knowledge base with three-tier structure is realized which makes a certain amount of complex natural language sentences to be converted into more instances with the teaching value.