

<<英语语料库与自动语法分析>>

图书基本信息

书名：<<英语语料库与自动语法分析>>

13位ISBN编号：9787100056595

10位ISBN编号：7100056594

出版时间：2007-11

出版单位：商务印书馆

作者：方称宇

页数：225

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

<<英语语料库与自动语法分析>>

内容概要

语料库语言学和计算语言学为促进自然语言处理技术快速发展的两门基础学科。

《英语语料库与自动语法分析》系这两个领域的一本专著，它以国际英语语料库为背景，着重探讨大型语料库的语法分析，尤其是英语口语材料给计算机自动处理带来的一系列难题。

书中涉及基于概率的自动词类识别和基于实例的自动句法分析这两大技术，并有专门章节来探讨句法分析的评测问题，对AUTASYS和The Survey Parser这两个软件系统的实际表现进行了深入的量化评测。此外，本书还探讨了介词短语的自动分析，特别是这类短语的句法功能的自动判定，并对自动语法分析在语音合成及语音识别中的应用做了相应的说明。

<<英语语料库与自动语法分析>>

书籍目录

Preface前言 List of Figures List of Tables Abstract

1. Introduction 1.1. What is Parsing? 1.2. The Introspective View 1.3. The Retrospective View 1.4. Data-Oriented Parsing 1.5. General Problems 1.6. The Proposed Research 1.6.1. Background to the Proposed Research 1.6.2. The Basic Approach of the Proposed Research 1.6.3. The Strengths and Novelties of the Proposed Approach 1.6.3.1. Automated Grammar Generation 1.6.3.2. De-Lexicalised Terminal Nodes 1.6.3.3. Global Parse with Subcategorisation Features 1.6.3.4. High-Quality Partial Parse 1.6.3.5. Intrinsic Ability to Learn 1.7. The Organisation of the Book

2. The Automatic Analysis of English Word Classes 2.1. An Overview of Word Class Tagging 2.2. Major Word Class Tagging Schemes 2.2.1. The Lancaster-Oslo/Bergen Tagging Scheme 2.2.1.1. The Lancaster-Oslo-Bergen Corpus 2.2.1.2. The Lancaster-Oslo-Bergen Tag Set 2.2.1.3. Summary 2.2.2. The International Corpus of English Tagging Scheme 2.2.2.1. The International Corpus of English 2.2.2.2. The International Corpus of English Tag Set 2.2.3. A Comparison of LOB and ICE 2.3. Word Class Tagging Methodologies 2.3.1. The Rule-Based Approach 2.3.2. The Probabilistic Approach

2.4. AUTASYS: A Hybrid Tagging System 2.4.1. A Probabilistic Approach Using the LOB Tag Set 2.4.1.1. The Tag Assignment Module 2.4.1.1.1. Tokenisation 2.4.1.1.2. The treatment of "." 2.4.1.1.3. The treatment of "" 2.4.1.1.4. Sentence boundary markers 2.4.1.2. Orthographic Analysis 2.4.1.3. Lexicon Lookup 2.4.1.3.1. The lexicon 2.4.1.3.2. The coverage of the lexicon 2.4.1.4. Morphological Analysis 2.4.2. The Idiom Identification Module 2.4.3. The Probabilistic Tag Selection Module 2.4.3.1. The Bigram Probabilistic Matrix 2.4.3.2. Implementing Probabilistic Tag Selection 2.4.4. The Rule-Based Refinement Module 2.4.5. Empirical Evaluation 2.4.6. Permissive AUTASYS-LOB Disagreements 2.4.6.1. NNP-NPT 2.4.6.2. JJ-JJB 2.4.6.3. NNP-NPL 2.4.6.4. RB-NN 2.4.7. Summary 2.5. A Rule-Based Approach towards LOB to ICE Translation 2.5.1. Solutions for Verbs 2.5.1.1. Auxiliary vs. Lexical 2.5.1.2. Monotransitive vs. Complex Transitive 2.5.1.3. Finite vs. Nonfinite 2.5.2. Closed Sets 2.5.3. Initial Results 2.5.4. Problems 2.5.5. Summary

3. The Automatic Induction of a Formal Grammar 4. Robust Practical Analogy-Based Parsing 5. Extensive Evaluations of the Survey Parser 6. The Resolution of Prepositional Phrases 7. Conclusions and Further Work References

Appendix A: A List of LOB Tags Appendix B: A List of ICE Tags Appendix C: A List of AUTASYS Idioms Appendix D: A List of ICE Parsing Symbols Appendix E: A List of ICE Prepositions in Descending Frequency Order Appendix F: A Distributional Profile of ICE-GB Prepositions Index

编辑推荐

本书的主要思路就是将已经分析过的语料库变成一个句法知识库，从中提取短语结构语法规则，并通过基于实例的手段，在知识库中为待分析语句提取一棵最佳句法树。本书对上述各个部分的研究进行了详细的描述，对系统的实际表现进行了深入的量化评测，并有专门章节来探讨句法分析的评测问题。除此之外，还探讨了介词短语的自动分析，特别是这类短语的句法功能的自动判定，因为这一研究和句法相似度分析有着密切的关系。同时，本书还就自动语法分析在语音合成及语音识别中的应用做了相应的介绍和说明，希望对读者能有所帮助。

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>