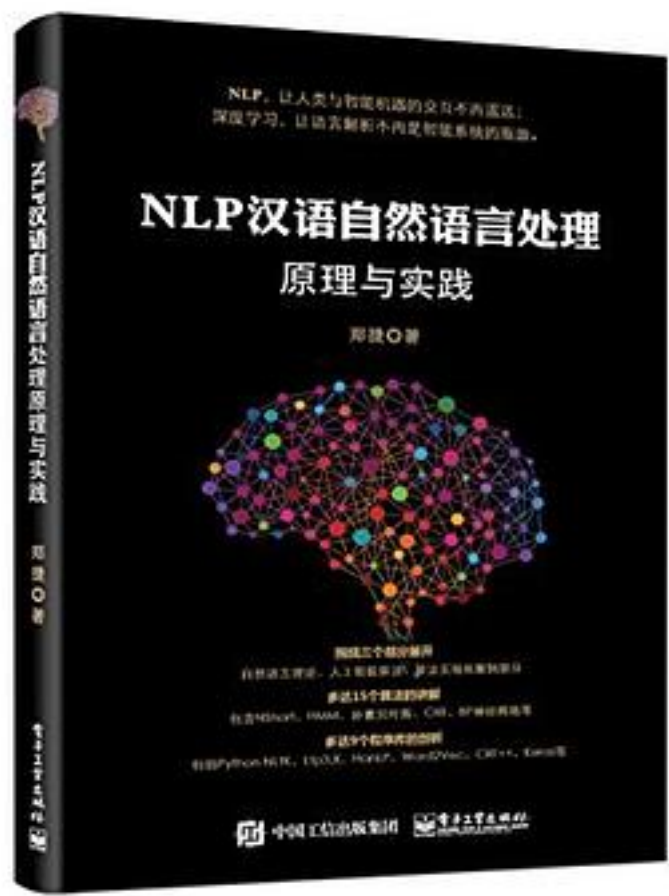


# NLP汉语自然语言处理原理与实践



[NLP汉语自然语言处理原理与实践\\_下载链接1](#)

著者:郑捷

出版者:电子工业出版社

出版时间:2017-1-1

装帧:平装

isbn:9787121307652

本书是一本研究汉语自然语言处理方面的基础性、综合性书籍，涉及NLP的语言理论、算法和工程实践的方方面面，内容繁杂。  
本书包括NLP的语言理论部分、算法部分、案例部分，涉及汉语的发展历史、传统的句法理论、认知语言学理论。需要指出的是，本书是迄今为止第一本系统介绍认知语言学和算法设计相结合的中文NLP书籍，并从认知语言学的视角重新认识和分析了NLP的句法和语义相结合的数据结构。这也是本书的创新之处。  
本书适用于所有想学习NLP的技术人员，包括各大人工智能实验室、软件学院等专业机构。

## 作者简介:

郑捷：网站的负责人，研究方向是机器学习与自然语言处理。当前负责的核心产品是高精度自然语言认知系统的设计与研发，研发目标是高精度（识别率为85%~95%）的统一架构的NLP认知系统，已经出版专著《机器学习算法原理与编程实践》，希望能与在NLP这方面有兴趣的读者一起学习交流。

目录: 第1章 中文语言的机器处理	1
1.1 历史回顾	2
1.1.1 从科幻到现实	2
1.1.2 早期的探索	3
1.1.3 规则派还是统计派	3
1.1.4 从机器学习到认知计算	5
1.2 现代自然语言系统简介	6
1.2.1 NLP流程与开源框架	6
1.2.2 哈工大NLP平台及其演示环境	9
1.2.3 Stanford NLP团队及其演示环境	11
1.2.4 NLTK开发环境	13
1.3 整合中文分词模块	16
1.3.1 安装Ltp Python组件	17
1.3.2 使用Ltp 3.3进行中文分词	18
1.3.3 使用结巴分词模块	20
1.4 整合词性标注模块	22
1.4.1 Ltp 3.3词性标注	23
1.4.2 安装StanfordNLP并编写Python接口类	24
1.4.3 执行Stanford词性标注	28
1.5 整合命名实体识别模块	29
1.5.1 Ltp 3.3命名实体识别	29
1.5.2 Stanford命名实体识别	30
1.6 整合句法解析模块	32
1.6.1 Ltp 3.3句法依存树	33
1.6.2 Stanford Parser类	35
1.6.3 Stanford短语结构树	36
1.6.4 Stanford依存句法树	37
1.7 整合语义角色标注模块	38
1.8 结语	40
第2章 汉语语言学研究回顾	42
2.1 文字符号的起源	42
2.1.1 从记事谈起	43
2.1.2 古文字的形成	47
2.2 六书及其他	48
2.2.1 象形	48
2.2.2 指事	50

2.2.3 会意	51
2.2.4 形声	53
2.2.5 转注	54
2.2.6 假借	55
2.3 字形的流变	56
2.3.1 笔与墨的形成与变革	56
2.3.2 隶变的方式	58
2.3.3 汉字的符号化与结构	61
2.4 汉语的发展	67
2.4.1 完整语义的基本形式——句子	68
2.4.2 语言的初始形态与文言文	71
2.4.3 白话文与复音词	73
2.4.4 白话文与句法研究	78
2.5 三个平面中的语义研究	80
2.5.1 词汇与本体论	81
2.5.2 格语法及其框架	84
2.6 结语	86
第3章 词汇与分词技术	88
3.1 中文分词	89
3.1.1 什么是词与分词规范	90
3.1.2 两种分词标准	93
3.1.3 歧义、机械分词、语言模型	94
3.1.4 词汇的构成与未登录词	97
3.2 系统总体流程与词典结构	98
3.2.1 概述	98
3.2.2 中文分词流程	99
3.2.3 分词词典结构	103
3.2.4 命名实体的词典结构	105
3.2.5 词典的存储结构	108
3.3 算法部分源码解析	111
3.3.1 系统配置	112
3.3.2 Main方法与例句	113
3.3.3 句子切分	113
3.3.4 分词流程	117
3.3.5 一元词网	118
3.3.6 二元词图	125
3.3.7 NShort算法原理	130
3.3.8 后处理规则集	136
3.3.9 命名实体识别	137
3.3.10 细分阶段与最短路径	140
3.4 结语	142
第4章 NLP中的概率图模型	143
4.1 概率论回顾	143
4.1.1 多元概率论的几个基本概念	144
4.1.2 贝叶斯与朴素贝叶斯算法	146
4.1.3 文本分类	148
4.1.4 文本分类的实现	151
4.2 信息熵	154
4.2.1 信息量与信息熵	154
4.2.2 互信息、联合熵、条件熵	156
4.2.3 交叉熵和KL散度	158
4.2.4 信息熵的NLP的意义	159
4.3 NLP与概率图模型	160
4.3.1 概率图模型的几个基本问题	161

4.3.2 产生式模型和判别式模型	162
4.3.3 统计语言模型与NLP算法设计	164
4.3.4 极大似然估计	167
4.4 隐马尔科夫模型简介	169
4.4.1 马尔科夫链	169
4.4.2 隐马尔科夫模型	170
4.4.3 HMMs的一个实例	171
4.4.4 Viterbi算法的实现	176
4.5 最大熵模型	179
4.5.1 从词性标注谈起	179
4.5.2 特征和约束	181
4.5.3 最大熵原理	183
4.5.4 公式推导	185
4.5.5 对偶问题的极大似然估计	186
4.5.6 GIS实现	188
4.6 条件随机场模型	193
4.6.1 随机场	193
4.6.2 无向图的团 (Clique) 与因子分解	194
4.6.3 线性链条件随机场	195
4.6.4 CRF的概率计算	198
4.6.5 CRF的参数学习	199
4.6.6 CRF预测标签	200
4.7 结语	201
第5章 词性、语块与命名实体识别	202
5.1 汉语词性标注	203
5.1.1 汉语的词性	203
5.1.2 宾州树库的词性标注规范	205
5.1.3 stanfordNLP标注词性	210
5.1.4 训练模型文件	213
5.2 语义组块标注	219
5.2.1 语义组块的种类	220
5.2.2 细说NP	221
5.2.3 细说VP	223
5.2.4 其他语义块	227
5.2.5 语义块的抽取	229
5.2.6 CRF的使用	232
5.3 命名实体识别	240
5.3.1 命名实体	241
5.3.2 分词架构与专名词典	243
5.3.3 算法的策略——词典与统计相结合	245
5.3.4 算法的策略——层叠式架构	252
5.4 结语	259
第6章 句法理论与自动分析	260
6.1 转换生成语法	261
6.1.1 乔姆斯基的语言观	261
6.1.2 短语结构文法	263
6.1.3 汉语句类	269
6.1.4 谓词论元与空范畴	274
6.1.5 轻动词分析理论	279
6.1.6 NLTK操作句法树	280
6.2 依存句法理论	283
6.2.1 配价理论	283
6.2.2 配价词典	285
6.2.3 依存理论概述	287

6.2.4 Ltp依存分析介绍	290
6.2.5 Stanford依存转换、解析	293
6.3 PCFG短语结构句法分析	298
6.3.1 PCFG短语结构	298
6.3.2 内向算法和外向算法	301
6.3.3 Viterbi算法	303
6.3.4 参数估计	304
6.3.5 Stanford 的PCFG算法训练	305
6.4 结语	310
第7章 建设语言资源库	311
7.1 语料库概述	311
7.1.1 语料库的简史	312
7.1.2 语言资源库的分类	314
7.1.3 语料库的设计实例：国家语委语料库	315
7.1.4 语料库的层次加工	321
7.2 语法语料库	323
7.2.1 中文分词语料库	323
7.2.2 中文分词的测评	326
7.2.3 宾州大学CTB简介	327
7.3 语义知识库	333
7.3.1 知识库与HowNet简介	333
7.3.2 发掘义原	334
7.3.3 语义角色	336
7.3.4 分类原则与事件分类	344
7.3.5 实体分类	347
7.3.6 属性与分类	352
7.3.7 相似度计算与实例	353
7.4 语义网与百科知识库	360
7.4.1 语义网理论介绍	360
7.4.2 维基百科知识库	364
7.4.3 DBpedia抽取原理	365
7.5 结语	368
第8章 语义与认知	370
8.1 回顾现代语义学	371
8.1.1 语义三角论	371
8.1.2 语义场论	373
8.1.3 基于逻辑的语义学	376
8.2 认知语言学概述	377
8.2.1 象似性原理	379
8.2.2 顺序象似性	380
8.2.3 距离象似性	380
8.2.4 重叠象似性	381
8.3 意象图式的构成	383
8.3.1 主观性与焦点	383
8.3.2 范畴化：概念的认知	385
8.3.3 主体与背景	390
8.3.4 意象图式	392
8.3.5 社交中的图式	396
8.3.6 完形：压缩与省略	398
8.4 隐喻与转喻	401
8.4.1 隐喻的结构	402
8.4.2 隐喻的认知本质	403
8.4.3 隐喻计算的系统架构	405
8.4.4 隐喻计算的实现	408

8.5 构式语法	412
8.5.1 构式的概念	413
8.5.2 句法与构式	415
8.5.3 构式知识库	417
8.6 结语	420
第9章 NLP中的深度学习	422
9.1 神经网络回顾	422
9.1.1 神经网络框架	423
9.1.2 梯度下降法推导	425
9.1.3 梯度下降法的实现	427
9.1.4 BP神经网络介绍和推导	430
9.2 Word2Vec简介	433
9.2.1 词向量及其表达	434
9.2.2 Word2Vec的算法原理	436
9.2.3 训练词向量	439
9.2.4 大规模上下位关系的自动识别	443
9.3 NLP与RNN	448
9.3.1 Simple-RNN	449
9.3.2 LSTM原理	454
9.3.3 LSTM的Python实现	460
9.4 深度学习框架与应用	467
9.4.1 Keras框架介绍	467
9.4.2 Keras序列标注	471
9.4.3 依存句法的算法原理	478
9.4.4 Stanford依存解析的训练过程	483
9.5 结语	488
第10章 语义计算的架构	490
10.1 句子的语义和语法预处理	490
10.1.1 长句切分和融合	491
10.1.2 共指消解	496
10.2 语义角色	502
10.2.1 谓词论元与语义角色	502
10.2.2 PropBank简介	505
10.2.3 CPB中的特殊句式	506
10.2.4 名词性谓词的语义角色	509
10.2.5 PropBank展开	512
10.3 句子的语义解析	517
10.3.1 语义依存	517
10.3.2 完整架构	524
10.3.3 实体关系抽取	527
10.4 结语	531
• • • • •	(收起)

[NLP汉语自然语言处理原理与实践 下载链接1](#)

标签

NLP

人工智能

自然语言处理

汉语

计算机

python

机器学习

语言学

## 评论

有些拿论文充书之嫌

-----  
这本是下错书的偶然收获，此NLP非彼NLP，对行业内或是浅白的科普，对行业外近乎天书，神经网络深度学习，两大阵营的造神运动似乎早已拉开了序幕，有生之年或许我们还会看到把AI当做信仰的宗教，规则还是统计？这是一个问题……

-----  
入门挺好的，起码想研究其中某一点，可以很快速的找到相关技术名称，然后百度/google。。(逐渐发现书中错误不少。。。。。。)

-----  
失望

-----  
没屁用

-----  
东拉西扯让人发笑

-----  
不错，可以入门简单了解

-----  
不行了，我收回觉得写得可以，错误不少，且存在链接失效

-----  
概率图模型那还行，相比我目前读到的其它书而言。

-----  
东拼西凑，数学概念讲的也不是很清楚，还时不时推销自己另一本书。

-----  
第二章讲汉语语言学历史？？？...整本书没什么干货 就当科普了 浪费钱 不推荐

-----  
我看完了这本厚书,写的很好的,  
全篇下来深深感觉到前后的行文逻辑的完整性.整本书应该是作者一人完成.  
每章的开头和结尾,是作者思维的发光点,  
告诉你起承转合,让我懂得了知识结构是怎样耦合的.

-----  
这本书比较渣，，讲真的，不咋样但是我还是会去读完，写上我的见解和笔记

-----  
东抄抄西抄抄。作为一个学习提纲还凑合

-----  
这应该是一篇论文或文章吧，本来是英文的，翻译成了中文。原文是用英语写的，用英文来介绍中文，所以翻译成中文之后，有些地方读起来怪怪的。作为一篇导论，很多算法没有具体讲，要想弄懂还是要去看其他的书吧。



-----  
中文NLP的教材看过不少，这一本 还凑合

-----  
[NLP汉语自然语言处理原理与实践\\_下载链接1](#)

书评

-----  
[NLP汉语自然语言处理原理与实践\\_下载链接1](#)