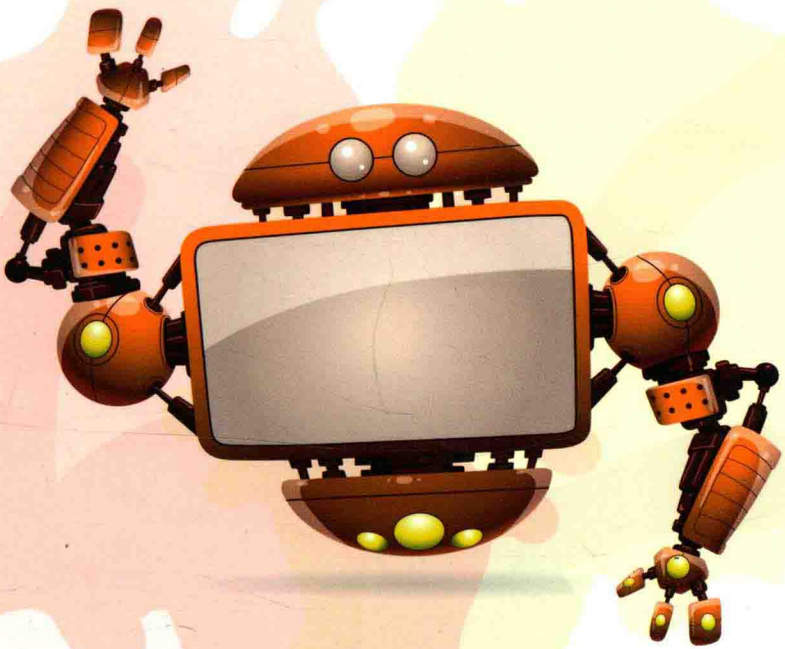
Broadview®
www.broadview.com.cn



自然语言处理

原理与技术实现

罗刚　张子宪 ／编著

中国工信出版集团
電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

自然语言处理

原理与技术实现

罗刚　张子宪 ／编著

電子工業出版社
Publishing House of Electronics Industry
北京•BEIJING

内 容 简 介

自然语言处理技术已经深入我们的日常生活。我们经常用到的搜索引擎就用到了自然语言理解等自然语言处理技术。自然语言处理是一门交叉学科，涉及计算机、数学、语言学等领域的知识。

本书详细介绍中文和英文自然语言处理的原理，并以 Java 实现，包括中文分词、词性标注、依存句法分析等。其中详细介绍了中文分词和词性标注的过程及相关算法，如隐马尔可夫模型等。在自然语言处理的应用领域主要介绍了信息抽取、自动文摘、文本分类等领域的基本理论和实现过程，此外还有问答系统、语音识别等目前应用非常广泛的领域。在问答系统的介绍中，本书特地介绍了聊天机器人的实现过程，从句子理解、句法分析、同义词提取等方面揭示聊天机器人的实现原理。

本书详细介绍自然语言处理的各个领域，既有理论，也有实现过程。对于打算从事自然语言处理研究的计算机、数学或语言学领域的专业人士，本书是难得的入门教材。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目（CIP）数据

自然语言处理原理与技术实现 / 罗刚，张子宪编著. —北京：电子工业出版社，2016.5
ISBN 978-7-121-28620-9

Ⅰ. ①自… Ⅱ. ①罗… ②张… Ⅲ. ①自然语言处理 Ⅳ. ①TP391

中国版本图书馆 CIP 数据核字（2016）第 082131 号

责任编辑：董　英
印　　刷：中国电影出版社印刷厂
装　　订：三河市华成印务有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　　邮编：100036
开　　本：787×980　1/16　　印张：27.75　　字数：618 千字
版　　次：2016 年 5 月第 1 版
印　　次：2016 年 5 月第 1 次印刷
印　　数：3000 册　　定价：79.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：010-51260888-819　faq@phei.com.cn。

前言

目前，互联网上的信息迅速膨胀，要想从中找出需要的信息就需要使用搜索引擎，你是否知道搜索引擎的工作原理？它到底如何对诸如汉语这样的文字进行加工、然后反馈出需要的结果？像这样的语言处理问题都会在本书中找到答案。即使在计算机专业，也有很多人对这个学科很陌生。因此，本书借助流行的 Java 语言介绍自然语言处理的各个领域，希望为推动相关应用的发展做出贡献。

本书的编者在自然语言处理教学和软件开发过程中积累的经验融入到本书的各个环节，读者会因此感到原理和实际应用结合得非常紧密。编者的实践经验还体现在相关的其他书中，如《自己动手写搜索引擎》《自己动手写网络爬虫》《自己动手写网络爬虫》《使用 C#开发搜索引擎》《解密搜索引擎技术实战》等。

有一些自然语言处理的开发原理与技巧在专业的公司内部秘而不宣。理论与实践结合的专门讲自然语言处理的书籍仍然相对较少。本书尝试先介绍原理，接着是具体的代码实现分析。本书相关的代码在读者 QQ 群（499526946）中的共享文件中可以找到。

国外有的基础课程从“构建搜索引擎”开始介绍计算机科学。我们的相关培训课程已经这样做了。当年参加过培训的学员，有些已经创业成功或者成为公司的技术骨干。要根据培训经验写出很好的自学教程，我们还有很多工作要做。零基础自学的读者，可能还需要其他方式来补足。

自然语言处理开发岗位比较少。如果能够花若干年开写出自己的软件产品，那么就可以合

伙创业了。这个过程对很多人来说，往往太漫长。像老外那样把冰箱放满匹萨饼和可乐，然后就开始干活，这样往往行不通，因为那样吃不了几天。可以买好能够保存几十年的谷子、水培可以吃若干年的韭菜。

就好像放在水里的韭菜种子，刚开始几天根本看不到变化，学习是个循序渐进的过程。可以在读者群中共同学习。

感谢开源软件和我们的家人，关心我们的老师和朋友们、创业伙伴，以及选择猎兔自然语言处理软件的客户多年来的支持。

特别提醒大家：经常面对电脑，容易阻塞气血。往往并没有免费的程序员保健师帮忙，所以需要自己多压腿，拉伸身体。多做腹部运动，减少腹部脂肪堆积，避免脂肪肝等疾病。此外，还可以拍打身体，例如腋下、臂弯、腘窝等关节凹下去的地方。

长时间对着散发蓝光的电脑屏幕容易失眠。为了提高睡眠质量，可以经常吃小米、藕、虾皮、鸡蛋等，喝决明子、玉兰花、熏衣草、绞股蓝等花草茶。

参与本书编写的还有石天盈、张进威、刘宇、张继红、徐友峰、何淑琴、孙宽、任通通、高丹丹，特别鸣谢！

目录

欢迎访问：电子书学习和下载网站（https://www.shgis.com）
文档名称：《自然语言处理原理与技术实现》罗刚，张子宪 编著.pdf
请登录 https://shgis.com/post/3428.html 下载完整文档。
手机端请扫码查看: