

<<知识工程语言学>>

图书基本信息

书名：<<知识工程语言学>>

13位ISBN编号：9787302222156

10位ISBN编号：7302222150

出版时间：2010-6

出版时间：清华大学出版社

作者：鲁川

页数：357

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>

<<知识工程语言学>>

前言

我要告诉读者：为什么要写这本书？

这本书是怎样写成的？

书中主要写了些什么？

半个世纪前的1956年初夏，中国大地上响彻了“向科学进军”的号角声，那时我正要填报高考志愿。在所能阅读到的《知识就是力量》等科普刊物中，最令我神往的就是那刚诞生了10年的“电子计算机”。

那一年招计算机专业的高等院校只有一所，就是哈尔滨工业大学。

我幸运考取之后，产生了一个想法：既然“机器人”是以“计算机”为大脑的，那么这个大脑就不应该只会“计算”，还应懂得“语言”、会做“文章”。

从上大学一年级开始我就在学好“高等数学”等规定课程的同时，找来中文系的各种课本默默地自学起来。

后来才知道恰在我入学的1956年暑期，在美国的达特茅斯（Dartmouth）召开的学术会议宣告了“人工智能”的诞生，也知道了新兴的“自然语言理解”等研究领域。

这坚定了我对“计算机能理解人类语言”的信心，也增强了“我应该自学语言学”的恒心。

我在世界“人工智能”诞生的1956年暑期考上计算机系，这个巧合是我的幸运，也决定了我一生的命运。

1961年从计算机系毕业以后的20年间，我为了“人工智能”的目的，坚持自学古代汉语、现代汉语、理论语言学和外语，注意追踪国外语言学主要学派的最新成果。

1982年我慕名拜访了范继淹先生，他是中国社会科学院语言研究所著名的语言学家、中国人工智能学会的副理事长、我国“计算语言学”的开拓者之一。

我向范先生表达了要报考他的“计算语言学”研究生的愿望。

范先生告诉我：超过规定报考年龄的人是不准报考的。

在我的恳切请求下，范先生对我进行了口试，证实了我的语言学基础知识远远超过了研究生的入学水平，答应向领导请示可否破例特招，结果没有批准。

但范先生已被我的特殊经历和优异成绩打动，决定破例招收我为“不发文凭的在职研究生”，悉心辅导我学习“计算语言学硕士研究生”的全部课程。

拜范先生为师之后才听别人说，早在1973年他就被确诊患了血癌，为了给国家培养急需的计算语言学人才，他分秒必争地跟死神抢夺时间而忍受着难以想象的痛苦。

1985年春天，范先生正式宣布我达到了“计算语言学硕士研究生”的结业水平。

不久范先生就住进了医院，在病床上恩师对我说了他的临终嘱托：恩师已经在给《中国语文》

（1985年第5期）的稿件《无定NP主语句》中提到我的名字，在参考文献中引用了我的一篇会议论文，目的是引起语言学界的关注。

让我尽早在语言学期刊上发表论文以争取语言学界的指导。

前几年恩师跟陆俭明和邢福义先生等中年语言学家发起的、有语言学大师吕叔湘和朱德熙先生出席的“现代汉语语法学术讨论会”已召开了三次。

范先生知道自己不可能参加第四次讨论会了，他跟会议负责人说过，要我接替他参加讨论会，让我赶快写一篇论文在会上宣读。

让我全力筹建“计算语言学”的全国性学术团体。

<<知识工程语言学>>

内容概要

本书是作者在为中国科学院研究生院和中国邮电大学研究生讲授“知识工程语言学”课程时所写的讲义的基础上完成的。

这本书为人工智能、知识工程及相关专业的人员提供了所需的最必要的语言学基础理论和最有用的语言数据，包含对汉语特色的最新认识以及跟英语的对比。

它凝聚了作者数十年从事知识工程领域科研工作和坚持不懈钻研语言学的心得。

本书可供从事人工智能、知识工程、自然语言理解、中文信息处理、机器翻译的研究人员参考，也可供从事对外汉语教学、汉语国际推广、英汉对比研究的教师和研究者参考，还可以作为高等院校相关专业的高年级本科生和研究生的选修课教材或参考书。

作者简介

鲁川，河南开封市人，1961年毕业于哈尔滨工业大学计算机系。

长期从事中文信息处理和计算语言学的研究和教学工作。

曾任北京信息工程学院教授，教育部语言文字应用研究所计算语言学研究室研究员，中国中文信息学会计算语言学专业委员会首届主任，中国人工智能学会自然语言理解专业委员会委员兼机器学习专业委员会委员，中国应用语言学会理事，《中文信息学报》编委。

研究领域为人工智能、知识工程、自然语言理解、机器翻译、英汉对比、对外汉语教学等。

出版的专著有《汉语语法的意合网络》。

《Learn Chinese Through 100 Sentence Frames》等，主编了《动词大词典(人机两用)》等。

在《中国语言学报》、《汉语学习》、《语言教学与研究》、《世界汉语教学》、《语言科学》等核心期刊上发表过多篇论文。

作为项目负责人，多次主持完成了国家级科研项目，如国家语言文字应用科研项目“信息处理用规范汉字字义统计和造词模式”和国家社会科学基金项目“信息处理与对外汉语教学的句子语序模式”等。

所设计的从汉字输入技术到对外汉语教学一条龙的“汉语自学机”获得了国家发明专利。

<<知识工程语言学>>

书籍目录

上篇 导论 第1章 信息时代需要知识工程语言学	1.1 信息	1.1.1 信息是物质的基本属性之	1.1.2 信息的定义	1.1.3 人类文明的三个时代	1.1.4 信息时代的高级阶段必然
出现知识经济	1.2 知识	1.2.1 信息有待于优化和系统化	1.2.2 知识的定义	1.2.3 知识的层次	1.3 智能
知识的层次	1.3 智能	1.3.1 知识处理主要包括知识获取、知识传播和知识运用	1.3.2 智能的定义	1.4 人工智能	1.4.1 人工智能的诞生和初期的发展
自然语言理解应该紧密结合	1.5 知识工程	1.5.1 知识工程的提出	1.5.2 初期的知识工程主要是专家系统	1.6 自然语言理解	1.6.1 自然语言的特点
自然语言理解应该紧密结合	1.6 自然语言理解	1.6.1 自然语言的特点	1.6.2 自然语言理解的难点及其原因	1.6.3 自然语言的自释性和突破其理解难点的方法	1.6.4 自然语言理解的进展拓宽了知识工程的范围
自然语言理解应该紧密结合	1.7 知识工程语言学	1.7.1 新兴的作为交叉学科的语言学分支的融合	1.7.2 知识工程语言学的研究内容	第2章 语言是知识的编码系统	2.1 人类语言观的发展
自然语言理解应该紧密结合	2.1.1 语言是人类跟其他动物的主要区别之	2.1.2 语言是人类和计算机传递信息及实施控制的符号系统	2.2 语言是人类认识世界和表述知识的编码系统	2.2.1 客观世界·认知世界·语言世界	2.2.2 语言信息的发送和接收
自然语言理解应该紧密结合	2.3 语言的基本结构单位	2.3.1 汉语和英语的基本结构单位	2.3.2 汉语和英语基本结构单位的差异	2.4 认知模式对语言的制约	2.4.1 认知模式对语言基本结构单位的制约
自然语言理解应该紧密结合	2.4.1 认知模式对语言基本结构单位的制约	2.4.2 认知模式对语言基本结构顺序的制约	2.4.3 认知模式对语言类型的制约	2.5 语言研究的观察视角和剖析平面	2.5.1 语言研究的观察视角%
自然语言理解应该紧密结合	2.5.2 句子构件的剖析平面	2.6 语言的优化发展和人类的国际通用语	2.6.1 人类的语言正在逐步优化而发生重大变化	2.6.2 人类文明史上最成功的通用符号系统	2.6.3 衡量语言“科学性”的标准
自然语言理解应该紧密结合	2.6.4 推荐汉语作为国际通用语的候选者	2.6.5 国际通用语的基本条件	2.6.6 汉语要持续优化才可能成为国际通用语	2.6.7 “汉语一千字”成为通用语义符号的可行性	中篇 语言的库存单位
自然语言理解应该紧密结合	第3章 语形学：语言的光波载体和视觉感知	3.1 人类的刻写能力和文字的不同来源	3.1.1 人类的刻写能力是创造文字的原动力	3.1.2 汉语的文字是注重视觉信息的自源性文字	3.1.3 英语的文字是注重听觉信息的他源性文字
自然语言理解应该紧密结合	3.2 英语的拼写形式跟实际读音的关系	3.2.1 英语拼写跟读音关系复杂的原因	3.2.2 英语单个元音字母在四种音节类型中的读音	3.3 汉字的字形演变	3.3.1 汉字字形演变的主要阶段
自然语言理解应该紧密结合	3.3.2 汉字的简化和规范汉字	3.3.3 现代通用汉字印刷体的字号和字体 下篇 语言的交际单位	参考文献	

<<知识工程语言学>>

章节摘录

插图：“信息论”奠基人仙农（C.shannon）说：“信息是用以消除随机不确定性的东西。”应该说这个定义是个实用的定义。

尽管它说明的是信息的作用而没有说明信息的本身到底是什么，但这个定义明确指出了信息对人类的重要性。

人类在认识和改造客观世界的实践中必然遇到大量令人困惑的“不确定性”，正是依靠所获得的信息来消除了这些“不确定性”，取得了令人满意的成果。

（2）在各位学者研究成果基础上所建议的定义信息是物质的基本属性之一，是物质的间接存在性的自身显示，所显示的是该物质的存在状况。

信息的作用是能够消除不确定性。

这个定义阐明了“信息”的本质和作用。

“材质”和“能量”体现的是物质的“直接存在性”；而“信息”体现的是物质的“间接存在性”。

“间接存在性”显示“直接存在性”的状况。

例如，有三个塑料的英文字母块“E、N、T”。

从材质的角度看，无论其排列的状况如何，这三个塑料块的总质量是恒定的；从能量的角度看，如果将其静止地摆在同一地点，无论排列状况如何，这三个塑料块的总能量也是恒定的。

但是当排列顺序是“TEN”时，显示的信息是“十（ten）”；当排列顺序是“NET”时，显示的信息是“网（net）”。

“材质、能量、信息”一起，成为物质的三个基本属性。

基于这种认识，人们说当代前沿科学的三大支柱是“材料科学、能源科学、信息科学”。

“信息”都是自身显示的。

并非因被其他认识主体所反映才存在，而是在被反映之前就已经存在了。

“信息”都是物质自身发出的“自显信息”。

“自显信息”分为“所感信息”和“未感信息”。

因为人们能够研究的仅仅是“所感信息”，所以在不致误解的条件下，我们把“所感信息”简称为“信息”。

“信息”的作用是消除“不确定性”。

所以“信息”对人们来说是特别珍贵的。

考察一下人们“传递信息（说话）”的过程：人们在说话时，通常会首先说出一个“话题”（topic），然后再对这个话题加以“说明”（comment）。

这就是说，说话者先用“话题”来提出一个不确定的并引起听话者注意的悬念，然后再用“说明”去消除其中的不确定性，从而解除听话者的悬念。

1.1.3 人类文明的三个时代人类文明的三个时代是按照物质生产的发展水平来划分的。

人类生存于地球这个物质世界之中。

在“前文明时代”，人类跟地球上的一般动物一样，赖以生存的主要方式是觅食，即寻觅、采摘、捕捉、猎取大自然提供的现成的食物。

所以，前文明时代也称为“渔猎时代”。

文明时代的标志是人类运用自己创造的劳动工具从事物质生产，即生产包括食物在内的“生活资料”和包括工具在内的“生产资料”等。

人类文明时代按生产力的发展水平来分期。

即按照依次开发“材质资源、能量资源、信息资源”的顺序而分为下列三个时代。

<<知识工程语言学>>

编辑推荐

《知识工程语言学》是由清华大学出版社出版的。

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>