

苗传江 著



HNC (概念层次网络) 理论 **导论**

清华大学出版社



2

HNC（概念层次网络）理论导论

苗传江 著

清华大学出版社
北京

内 容 简 介

HNC 是英文 Hierarchical Network of Concepts(概念层次网络)的缩写,HNC 理论是一个关于自然语言理解处理的理论体系,它以概念联想脉络为主线,建立了一种表述和处理自然语言的新模式,这一模式不但能应用于自然语言的计算机理解处理,如机器翻译、智能检索、自动文摘、人机对话等领域,而且对自然语言研究的其他许多方面,如语言本体研究、语言认知研究等,都富有启发意义。HNC 是由中国人创立的新的语言理论,问世 10 年来,逐渐得到学界的普遍关注和认同,而且它的理论和技术已经在语言研究和语言信息处理系统及产品的研发中得到广泛的应用。

本书以 HNC 的创立者黄曾阳先生的专著《HNC(概念层次网络)理论——计算机理解语言研究的新思路》为基础,对 HNC 在理论、技术和知识库建设三个方面的要点与特色进行了全面而系统的阐释,是学习、研究和应用 HNC 理论的教程性著作。

本书适用于自然语言理解、语言学、语言信息处理、人工智能、认知科学等专业领域的研究者、开发者和学习者,以及其他对自然语言方面的科学问题感兴趣的读者。

版权所有,翻印必究。举报电话:010-62782989 13901104297 13801310933

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

本书防伪标签采用清华大学核研院专有核径迹膜防伪技术,用户可通过在图案表面涂抹清水,图案消失,水干后图案复现;或将表面膜揭下,放在白纸上用彩笔涂抹,图案在白纸上再现的方法识别真伪。

图书在版编目(CIP)数据

HNC(概念层次网络)理论导论/苗传江著. —北京:清华大学出版社,2005.1

ISBN 7-302-09868-9

I. H… II. 苗… III. 自然语言处理—研究 IV. ①H087②TP391

中国版本图书馆 CIP 数据核字(2004)第 113670 号

出 版 者: 清华大学出版社

地 址: 北京清华大学学研大厦

<http://www.tup.com.cn>

邮 编: 100084

社 总 机: 010-62770175

客户服务: 010 62776969

责任编辑: 赵彤伟

封面设计: 傅瑞学

版式设计: 肖 米

印 刷 者: 北京四季青印刷厂

装 订 者: 三河市新茂装订有限公司

发 行 者: 新华书店总店北京发行所

开 本: 188×230 印张: 22.5 字数: 459 千字

版 次: 2005 年 1 月第 1 版 2005 年 1 月第 1 次印刷

书 号: ISBN 7-302-09868-9/TP·6804

印 数: 1~1000

定 价: 68.00 元

本书如存在文字不清、漏印以及缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话:(010)62770175-3103 或(010)62795704

HNC 序

阅传江《HNC(概念层次网络)理论导论》待出版稿,序如下。

HNC 的第一本专著 1998 年出版以后不久,陈力为先生和林杏光先生都清晰而强烈地意识到必须再出版一本 HNC 的教程性著作。两位先生还不约而同地认为传江(当时还是在读博士,下文将简称苗博士或作者)是承担这一工作的最佳人选,于是多次鼓励苗博士勇挑这一重任并面授机宜。苗博士没有辜负两位先生的厚望,经过 4 年的努力,这本书终于问世了。两位先生的在天之灵一定深感欣慰。

《HNC(概念层次网络)理论导论》全面、系统地阐释了 HNC 理论、技术和知识库建设三方面的要点与特色,完全符合 HNC 教程的要求。达到这一要求是有很难度的,因为这一教程性的“导论”不是一般性的文献综述,需要创造性的再思考与实践。本书既是理论再思考的结晶,又是工程实践理论提升的样板。本书的上篇带有理论再思考的鲜明印记,读者可以从中清晰地看到 HNC 的全貌。其中关于特征语义块核心 Ek 复合构成的论述和广义对象语义块构成必须引入内容基元 C 的论述尤为突出。本书的下篇是 HNC 工程实践理论提升的生动体现,它是作者博士论文的扩展,读者从中可以感受到 HNC 知识观的魅力。

今年是 HNC 理论体系基本构架形成后的第一个十年,也是 HNC 第一本专著出版后的第一个五年。希望在随后的三个五年里依次再出版下列三部理论专著:《HNC 概念基元符号体系手册》、《句类知识手册》和《语境知识手册》。本书是第二部《手册》的先导,作者及 HNC 领域中的其他年轻才俊都应该为这三部《手册》的撰写作出自己的应有贡献。至于这三部《手册》对于自然语言理解获得最终突破的意义,若读者细心玩索此书,必能心领神会。

十五年太长了么?否!理论探索固然需要跟随市场的需求,但同样需要保持独立发展的品格,否则就会断送重大探索的生命。探索通常是不能设定时限的,但上述三部《手册》的探索由于已具备比较深厚的基础,故可以设定上

述时限。

HNC 是一个开放的、发展中的、通过不断自我反思以求得不断创新的理论和技术体系,如果说本书有什么弱点的话,那就是对 HNC 的这三项基本特征的描述着墨太少。这当然属于苛求,但是,像自然语言理解处理这样重大的理论探索是必须苛求的,否则就可能陷于满足现状而不知自拔的可悲状态。愿本书作者和读者牢记斯言。

黄曾阳

2003 年 8 月

前言

HNC(hierarchical network of concepts,概念层次网络)理论是一个关于自然语言理解(natural language understanding, NLU)处理的理论体系,由中国科学院声学研究所黄曾阳先生创立。该理论于1994年初步形成后,先后在多项国家重大课题中得到应用和检验,并不断发展和完善,逐步实现了技术化。而且,HNC的理论和技術近年来已经在语言信息处理产品的研发中得到广泛的应用。

HNC理论由语言概念空间考察自然语言空间,以概念联想脉络为主线,建立了一种表述和处理自然语言的新模式,这一模式不但能应用于自然语言的计算机理解处理,如机器翻译、智能检索、自动文摘、人机对话等领域,而且对自然语言研究的其他许多方面,如语言本体研究、语言认知研究等,都富有启发意义。因此,该理论得到越来越多的专家学者的关注和认同。

1998年,清华大学出版社出版了黄曾阳先生的专著《HNC(概念层次网络)理论——计算机理解语言研究的新思路》。这本专著出版以后,读者普遍认为HNC理论确实提供了一种新思路。同时,有不少读者反映这本书不太好懂,尽管这里面有种种不同的原因,但无论如何,写一本书来对HNC理论的基本内容进行相对简明、通俗的讲解和阐释是很有必要的。为该专著题词的陈力为院士和写“编者的话”的林杏光教授都希望我能尽快写出这样一本书来。时光飞逝,转眼就过了四年多,二老已先后逝世,我这本书现在才完稿,想来深感汗颜和遗憾。

我是在1996年底初次接触HNC的。当时,在导师林杏光先生的介绍下,我认识了黄先生,开始学习HNC理论。从那以后,我就一直在从事HNC的研究和开发工作,并且成为黄先生的博士生。HNC理论深深地吸引了我,我认为用它来进行语言研究和处理是大有可为的,几年来的实践已经证明了这一点。因此,我很想把这一理论介绍给更多的人,希望有更多的人来了解、学习、研究和应用它。

我曾在“HNC联合攻关组”内外多次介绍或讲解HNC理论。对我而言,

每一次都是很好的学习机会,其间所讨论的问题令我受益匪浅,它们促进和启发我加深对 HNC 的理解,也成为我写作本书的推动力。在“HNC 联合攻关组”内,我办过五六次讲习班。讲习班的目的是让进入攻关组的新人尽快熟悉 HNC 的基本内容,并能在这一理论体系下开始研发工作。讲习班所需的时间大约为二十个小时,所用的基本材料就是本书的提纲。历次讲习班的效果都还是不错的,我很高兴地看到,讲习班起到了比较好入门作用,参加者能够在较短的时间内掌握 HNC 的基本思路和方法,这表明 HNC 理论并不难学。本书的内容就是在这些讲解的基础上,经过逐步积累和修改而成的。

本书的内容也是我在培养知识库人才的过程中积累起来的。在 HNC 的工程项目中,我负责知识库的建设工作。知识库建设的基本任务是运用 HNC 的理论和方法来分析和描述语言事实,揭示语言规律,形成计算机处理自然语言所需的知识。知识库建设要面对各种各样的、具体的、真实的语言现象和问题,工作量巨大,必须培养相当数量的专门人才。从事知识库工作的人需具备运用 HNC 研究语言的良好能力,几年来的工作表明,在本书的基础上学习和掌握 HNC 理论,有助于培养这种能力。

关于本书的内容,还需说明三点:①本书讲解的是 HNC 的基本内容,有不少地方没有展开。要深入了解 HNC,最好的途径是研读黄先生的原著。因此,本书上篇部分的每一章后都设有“导读”,以引导读者查阅黄先生对一些重要概念和问题的相关论述。在正文的讲述中,也尽量注明黄先生专著中相关论述所在的页码,以便查阅。②为行文方便,本书在引述黄先生专著中的内容时,大多没有采用直接引用的方式。③HNC 还处在不断发展之中,黄先生的专著出版以后,HNC 又在不少方面取得了新的进展或进行了更新,本书酌情吸收了这些新内容。为了便于前后接轨,书中加了必要的注释,以说明其间的不同。读者如果要获取 HNC 的最新资料,请访问网站 <http://www.hnecnlp.com>。

本书的重点是讲解和阐述 HNC 的理论框架与基本思路,而较少涉及 HNC 理论对具体问题的研究和分析,但读者可以从中领会到如何应用 HNC 的思路和方法去研究和处理自然语言,如果真能如此,本书的主要目的就算达到了。

本书对 HNC 理论的阐述是基于笔者的理解,难免有错误或不当之处,敬请读者不吝批评指正。

苗传江
2003 年 4 月

致谢

感谢导师黄曾阳先生和林杏光先生，他们是对我帮助和影响最大的人。

黄先生那深邃的思想、广博的知识、高尚的人格和潜心钻研、精益求精的治学态度一直深深吸引着我，令我敬佩。他所创立的 HNC 理论让我看到了自然语言理解处理这一领域的广阔天地和美好前景。想起六年前我刚开始学习 HNC 的时候，常常在周末到黄先生家里请他答疑，那种学习的愉悦真令人神往。对我来说，不论是聆听黄先生的讲解，还是拜读黄先生的著述，都是一种享受。黄先生的言行是我内心的鞭策力量。

林先生是黄先生在科研战线上的挚友。是林先生引导我进入自然语言处理这个领域，也是他介绍我认识黄先生和学习 HNC，并支持和鼓励我以 HNC 的发展为事业。正是在林先生的指导和启发下，我逐步加深了对 HNC 的理解和把握。林先生那种坚韧不拔的治学精神，令我钦敬；他那种待学生如爱子的慈爱之情，令我感动。多年来，林先生一直关心着我的成长，我从他那里不仅得到学术上的指导，还得到生活上的关照和为人处事方面的教诲。在这里，让我对林先生的不幸逝世致以深切的悼念。

感谢先辈陈力为院士大力支持和鼓励我写作本书。黄先生的专著出版后，陈老曾多次叫到我家里讨论 HNC 的一些问题。当时他已是八十多岁的高龄，身体多有不便，但仍然十分关注中文信息处理和自然语言理解的最新进展，对 HNC 给予公允、中肯的评价与支持，这种精神令我敬仰。

感谢著名语言学家、全国人大副委员长许嘉璐教授对我的关心与支持。早在十年前 HNC 刚刚问世的时候，许先生就以一位博通古今的语言学家的眼光，敏锐地洞察到它对语言学和科学的重要意义与价值，从多方面对 HNC 的发展予以支持。许先生殷切希望我能致力于 HNC 在语言研究和信息处理中的应用，努力成为跨接语言学和计算机科学的“两栖”型人才，做出有新意的研究成果，这是我写作本书的巨大推动力。

感谢中国中文信息学会对本书给予充分肯定，并大力帮助和支持出版。特别感谢学会秘书长曹右琦教授和北京语言大学应用语言研究所所长张普

教授,他们在百忙中抽出时间认真审阅了本书的初稿,提出了很多非常宝贵的修改意见。

感谢清华大学出版社为本书提供难得的出版机会。

感谢“HNC 联合攻关组”内的所有同事,这个团队为我提供了良好的学习环境和工作的机会。特别感谢晋耀红博士,在 HNC 的研发中,他负责软件,我负责知识库,我们是密不可分的搭档。

感谢 HNC 研究院的董事长陈小盟先生,他的运作使 HNC 理论和技术得以在产品开发中得到应用,这是培育本书的肥沃土壤。也感谢研究院的所有同事,他们是本书的第一批读者,提出了很多宝贵的修改意见。

最后,感谢我的妻子蔡颖清,感谢她与我一起在北京工作和生活,给予我理解、支持和照顾,让我安心投入自己喜欢的工作。也感谢我远在沂蒙老家的父母,他们给了我良好的家教,供养我上大学,一直支持和理解我的选择,让我懂得珍惜自己所经历和拥有的一切。

苗传江

2003 年 8 月 6 日

简明目录

序	III
前言	V
致谢	VII
绪论	1

上 篇

第 1 章 语义网络——概念基元和联想脉络	9
第 2 章 概念表述模式——局部联想脉络	28
第 3 章 语句表述模式——全局联想脉络	50
第 4 章 知识库建设——知识表示体系	89
第 5 章 句类分析系统——语言理解技术	125

下 篇

第 6 章 句类知识总论	161
第 7 章 作用句的句类知识	171
第 8 章 效应句的句类知识	198
第 9 章 过程句的句类知识	206
第 10 章 转移句的句类知识	212
第 11 章 关系句的句类知识	226
第 12 章 状态句的句类知识	231
第 13 章 判断句的句类知识	236
第 14 章 混合句的句类知识	248
第 15 章 句类知识的应用	259
附录(1~5)	268
索引	336
参考文献	340

目 录

序	Ⅲ
前言	V
致谢	Ⅶ

绪论	1
0.1 HNC 的目标	1
0.2 HNC 的进展和基本内容	2
0.3 本书的内容和安排	5

上 篇

第 1 章 语义网络——概念基元和联想脉络	9
1.1 抽象概念和具体概念	9
1.2 抽象概念的三大语义网络	10
1.2.1 基本概念语义网络	10
1.2.2 基元概念语义网络	11
1.2.3 逻辑概念语义网络	18
1.3 语义网络的基本特征	20
1.4 语义网络的设计思想及其来源	22
1.5 其他语义网络	24
1.5.1 综合概念	24
1.5.2 语习概念	25
1.5.3 物化的基本概念	25
1.5.4 人化的基本概念	25
1.5.5 基本物概念	26
小结	26

导读	27
第 2 章 概念表述模式——局部联想脉络	28
2.1 概念的一般表达式	28
2.2 五元组和概念矩阵	30
2.3 概念的高层、中层和底层	35
2.4 概念的对比性、对偶性和包含性(中层层次符号)	36
2.5 本体层和挂靠层	38
2.6 概念的组合结构	41
2.6.1 作用效应类组合	42
2.6.2 对象内容类组合	42
2.6.3 逻辑类组合	43
2.6.4 语法类组合	44
2.6.5 关于组合结构的补充说明	45
小结	48
导读	49
第 3 章 语句表述模式——全局联想脉络	50
3.1 句类	50
3.2 语义块	52
3.3 语义块是句类的函数	54
3.4 句类表示式和句类代码	55
3.4.1 基本句类	55
3.4.2 混合句类	56
3.4.3 复合句类	58
3.5 辅语义块及两可语义块	59
3.5.1 辅语义块	59
3.5.2 两可语义块	62
3.6 句类知识	63
3.6.1 语句格式知识	63
3.6.2 语义块构成知识	66
3.6.3 语义块分离	75
3.6.4 句蜕和块扩	79
3.6.5 语义块之间的概念关联知识	83

3.6.6 句类转换和语义块变换知识	84
3.7 再谈语义块和句类	86
3.7.1 语义块概念提出的必要性	86
3.7.2 语义块与句子成分	86
3.7.3 语义块与格	86
3.7.4 句类是语句的深层结构	87
小结	87
导读	88
第4章 知识库建设——知识表示体系	89
4.1 知识库建设的根本原则	89
4.2 HNC 知识库系统	91
4.3 概念知识库	92
4.3.1 概念节点知识库	92
4.3.2 句类知识库	100
4.3.3 概念关联知识库	100
4.4 汉语词语知识库	102
4.4.1 概念类别	102
4.4.2 HNC 符号	108
4.4.3 句类代码	112
4.4.4 语句格式知识	115
4.4.5 广义对象语义块知识	116
4.4.6 特征语义块知识	117
4.4.7 角色知识	118
4.4.8 其他知识项	119
4.5 知识库建设的若干问题	121
小结	123
导读	124
第5章 句类分析系统——语言理解技术	125
5.1 HNC 理解处理系统的基本框架	125
5.2 预处理	126
5.3 语义块感知和句类假设	127
5.3.1 语义块感知	128

5.3.2 句类假设	130
5.4 句类检验	132
5.5 语义块构成分析	133
5.6 句类分析系统的其他模块	134
5.6.1 隐知识揭示	134
5.6.2 语境生成	136
5.6.3 新词伪词辨识	138
5.6.4 短时记忆	139
5.7 句类分析例析	139
5.8 关于语义距离及其计算	150
5.9 句类分析系统的输出结果及其应用	154
小结	156
导读	157

下 篇

第6章 句类知识总论	161
6.1 关于句类知识的若干问题	161
6.1.1 句类知识的系统与层次	161
6.1.2 句类知识的语种个性	162
6.1.3 句类知识的获取	163
6.1.4 句类知识与句类分析	164
6.1.5 句类知识的应用	164
6.2 句类知识通则	164
6.2.1 关于语句格式	164
6.2.2 关于语义块构成	166
6.2.3 关于辅语义块的位置	168
6.2.4 关于要素句蜕	169
6.2.5 关于无特征语义块句类	170
第7章 作用句的句类知识	171
7.1 基本作用句	171
7.2 作用效应句	176
7.3 承受句	178

7.3.1 一般承受句	179
7.3.2 主动承受句	182
7.3.3 被动承受句	183
7.3.4 特殊承受句	184
7.4 反应句	185
7.4.1 一般反应句	187
7.4.2 主动反应句	190
7.4.3 被动反应句	191
7.4.4 后续反应句	191
7.4.5 作用反应句	191
7.5 免除句	193
7.6 约束句	195
第8章 效应句的句类知识	198
8.1 一般效应句和基本效应句	199
8.2 双对象效应句	203
第9章 过程句的句类知识	206
9.1 一般过程句和基本过程句	206
9.2 素描句	209
9.3 因果果因句	210
第10章 转移句的句类知识	212
10.1 基本转移句	213
10.2 接收句	216
10.3 传输句	218
10.4 自身转移句	219
10.5 交换句	221
10.6 替代句	223
10.7 变换句	225
第11章 关系句的句类知识	226
11.1 双向关系句	228
11.2 单向关系句	229

11.3 扩展关系句·····	230
第 12 章 状态句的句类知识 ·····	231
12.1 一般状态句和基本状态句·····	231
12.2 换位状态句·····	233
12.3 简明状态句·····	234
第 13 章 判断句的句类知识 ·····	236
13.1 一般判断句·····	236
13.2 基本判断句·····	239
13.2.1 是否判断句·····	239
13.2.2 存在判断句·····	242
13.2.3 比较判断句·····	243
13.2.4 简明势态句·····	246
第 14 章 混合句的句类知识 ·····	248
14.1 广义作用句的混合·····	249
14.1.1 信息转移—作用效应句·····	249
14.1.2 反应—信息转移句·····	250
14.1.3 关系—反应句·····	251
14.1.4 关系—作用句·····	252
14.1.5 作用—替代句·····	253
14.1.6 作用—特殊承受句·····	253
14.1.7 作用反应—后续反应句·····	254
14.2 广义效应句的混合·····	254
14.3 广义作用句与广义效应句的混合·····	255
14.3.1 效应—作用句·····	255
14.3.2 反应—状态句·····	256
14.3.3 作用—双对象效应句·····	257
14.3.4 信息转移—效应句·····	258
第 15 章 句类知识的应用 ·····	259
15.1 在多动词处理中的应用·····	259
15.2 在专名识别中的应用·····	265



附录 1 语义网络概念节点表	268
附录 2 基本句类代码和表示式	306
附录 3 语句格式代码和表示式	316
附录 4 基本句类和常见混合句类例句	319
附录 5 概念和语句表示符号一览	332
索引	336
参考文献	340