

# 一种世界常识性知识的形式化表述方法

缪建明<sup>1,2</sup>, 郝惠宁<sup>2</sup>, 张 全<sup>2</sup>

(1. 中国科学院研究生院, 北京 100039; 2. 中国科学院声学研究所, 北京 100080)

**摘要:** 如何形式化表述世界知识并提供给计算机使用, 是制约自然语言理解处理实现突破的核心问题之一。该文立足于 HNC 理论的整体框架, 对世界知识按领域进行分类, 为每一领域赋予领域句类表示式, 为句群处理提供知识支持。同时, 以国家治理与管理领域为例, 阐述领域句类表示式的设计步骤以及如何使用。

**关键词:** HNC 理论; 句群; 语境单元; 领域句类代码

## Formal Indicative Method on World Common Knowledge

MIAO Jianming<sup>1,2</sup>, HAO Huining<sup>2</sup>, ZHANG Quan<sup>2</sup>

(1. Graduate School, Chinese Academy of Sciences, Beijing 100039; 2. Institute of Acoustics, Chinese Academy of Sciences, Beijing 100080)

**【Abstract】**The problem that the boundless world knowledge formalized to indicate and to provide the computer to use is regarded as the one of the key problems which restrict natural language understanding processing to break through. This article bases on the HNC theory overall frame, carries on the world knowledge according to different domain classification, and has endue with the domain sentence category expression to each domain, loads them into computer ahead of time. The article takes the national government and the management as an example, introduces the design procedure of the domain sentence category expression, and elaborates on how to provide the computer for use.

**【Key words】**HNC theory; Sentence group; Sentence group unit; Domain sentence code

为使计算机表现出智能, 唯一的办法就是使它拥有并能够运用知识, 这一问题成为制约自然语言理解, 乃至整个人工智能研究实现突破的核心问题之一<sup>[1]</sup>。概念层次网络(HNC)理论认为, 知识可分为 3 类: 概念知识, 语言知识, 常识及专业知识。从重要性上看, 概念知识和语言知识居主要地位, 常识性知识居于次要地位。目前, HNC 理论针对概念知识及语言知识的特点建成了对应的知识库, 但常识性知识库仍没能建成, 究其根源来自于以往未曾找到合适的分类体系对其进行完整的知识划分。经过 10 多年的深入研究, HNC 理论得出“常识性知识具有一定领域”的结论。

自然语言处理进入句群一级处理后, 句群的形式化表述问题更加凸出。句群的理解对段落和篇章的理解具有至关重要的影响, 段落和篇章中语义信息的提取需要建立在句群理解的基础上。HNC 理论认为: 人类语言的表述必然具有一定的语境领域知识, 句群是围绕着这些特定领域展开的一组话语<sup>[2]</sup>。每一个特定领域都对应应该领域世界知识, 将先验化的形式化领域世界知识提供给计算机使用, 句群一级的理解将可获得重大突破。为此, 我们针对每一个领域的不同概念树, 设计了相对应的领域句类表示式, 建立了该领域概念树下的世界常识知识形式化表示框架, 实现语言空间的句群到语言概念空间的概念树之间的映射, 同时也完成了从句群到语境单元之间的过渡性处理的前期工作。

### 1 设计步骤

一个领域句类表示式的完成需要以下 3 个步骤:

- (1) 分析概念树节点延伸结构表示;
- (2) 提取领域知识的自然语言描述;
- (3) 具体设计领域句类表示式。

HNC 理论目前已经设计出概念完备的概念关联知识库,

共包含 108 棵概念数, 并且对概念树以 10 种领域进行了不同的分类(具体分类可见文献[2])。概念树及其延伸结构构成了对世界常识知识的划分。领域句类表示式设计第 1 步即是分析概念树的具体节点延伸结构, 通过人工专家的工作, 从中提取领域知识的自然语言描述内容, 在此基础上, 具体设计领域句类表示式。

为了详细了解设计领域句类表示式的各个具体步骤, 以下将以概念树国家治理与管理(a12)的节点延伸结构国家治理与管理(a12t=a)为例, 说明领域句类表示式是如何设计和配置世界常识知识的。

#### 1.1 概念树总体结构

国家治理与管理(a12)属于专业活动(a)中政治活动(a1)的一棵概念树, 用来描述政治活动中有关国家治理与管理方面所需要的常识性知识, 也简称治国。概念树 a12 设置了 4 项一级延伸概念, 分别为交织延伸 a12t=a、定向延伸 a123、定向延伸 a127 和并列延伸描述 a12\k=3, 其中 a12t=a 描述国家治理 a129 和国家管理 a12a 之间的交织性知识, a123 描述治国面临的几项基本关系, a127 描述治国的基本运作方式, a12\k=3 描述治国谋略的 3 个特殊侧面。根据节点设计的虚实设计原则, a12 节点应重点体现国家治理方面的世界性知识, 管理的共同性世界知识则放在普通管理 a018 节点中配置。a12 概念树的节点延伸结构如下:

**基金项目:** 国家“973”计划基金资助项目“自然语言理解的交互引擎研究”(2004CB318104); 中科院声学研究所知识创新工程资助项目“HNC 语言知识处理理论及技术”

**作者简介:** 缪建明(1977-), 男, 博士生, 主研方向: 自然语言理解; 郝惠宁, 工程师; 张 全, 研究员、博导

**收稿日期:** 2006-05-20 **E-mail:** mjm\_77@sina.com

a12:(t=a,3,7,k=3;  
 9t=a,ae2m,3(e2m,eam,n),7(m,e2m),\|e2m,\|2\*t=a,\|3k=4;  
 3e2m(e1n,3),3e217,3ea1d01,3ea2c01,3ne2n,347,713,727,\|e2m  
 e2n;  
 3ne2n3)

其中的第 1 个延伸概念，即 a12t=a 的延伸结构表示式如下：

a12t=a “治理与管理”  
 a129 “国家治理”  
 a129t=a “治理的两项永恒课题”  
 a1299 “开拓性治理”  
 a129a “整顿性治理”  
 a12a “国家管理”  
 a12ae2m “国家管理的两个基本侧面”  
 a12ae21 “横向管理”  
 a12ae22 “纵向管理”

本文仅给出 a12 的总体节点延伸结构和 a12t=a 的延伸结构表示式，其它延伸结构表示式可见文献[1]。

## 1.2 世界知识描述

专业活动都存在治理与管理这两个不同的侧面，治理问题在政治活动中特别突出，主要由最高领导人对国家整个大政方针的制定及执行来体现。从人类发展历史上看，治理对整个国家政治具有重大而深远的意义。如美国的罗斯福新政和中国的改革开放。国家本属于组织机构的范畴，但又不同于一般的组织机构，是一个政治、经济、文化等都需要综合考虑的、极为复杂的组织机构，这也就决定了治国需要具备组织机构的一般特性，同时它又有独特的世界知识。主要体现在：国家治理属于管理的战略层面，国家管理属于管理的战术层面。国家治理的延伸结构必须描述治理的两项永恒课题：开拓性治理与整顿性治理。组织机构纵横管理必然具有纵横交错特性，横向机构具有地域的分布性和实施的执行性，纵向机构具有地域的集中性和实施的决策性，二者用来描述管理的两个基本侧面，这一特性在治国中表现得最为明显，因此在节点的设计中重点体现出来，为此政治管理的延伸结构设计横向与纵向管理。

## 1.3 领域句类表示式设计

治国 a12 从宏观和整体的角度来看主要体现在国家治理 a129 和国家管理 a12a 两方面上，二者体现出国家统治者对治国谋略的不同阶段。不同阶段对于治国需要考虑的对象和内容都不同，在世界知识的形式化表述上也需要作不同的处理。对此需要设计不同的领域句类表示式，体现治国不同阶段的世界性常识性知识。以下分别从国家治理 a129 和国家管理 a12a 两方面对领域句类表示式的设计思路进行阐述。

国家治理 a129 的全局领域句类表示式设计如下：

SCD=SCD(a129)=Cn-1|RtC(ReC)D01-42(Y101-1/Y501-0)\*211J  
 其中各语义块的对应如下：

DA=p44e61d01pj2(国家最高领导人);YBCB=ppj2(具体国家)  
 YBCC(a1299)=a103e21(具体国家的对内政策);  
 YBCC(a129a)=ra11t(政权各组成机构)

国家治理不同于国家管理，体现的是国家最高领导人对整个国家大政方针的整体谋略思想。“Cn-1|”用来表示时间的阶段性，国家治理有一定的阶段性，同时国家治理也带有一定的目的性，当治理目的达到后，治理自然结束，这一特点在表达式中通过“Cn-1|RtC”来表示；治国谋略又不同于一般性管理，这其中需要谋略者 DA 对古今中外国家治理的综合考虑，其中涉及到对以往治理的借鉴和参照，“ReC”体现了这一参照借鉴，但大多数情况下这一借鉴过程往往不说，

更体现开创性智慧，ReC 的括号用来表示两可的情况；开拓性治理 a1299 与整顿性治理 a129a 体现治理的两种不同类型，开拓性治理必然产生出以往未有之事物，整顿性治理则主要完成对已有事物的调整，这一区别通过基本句类代码 Y101-1 和 Y501-0 体现，但两者面对的对象一致，对象的内容则不同，开拓性治理主要在对内政策上进行治理，它的治理对象为具体国家 YBCB 的对内政策 YBCC，而整顿性治理对象为具体国家 YBCB 的政权各组成机构 YBCC。

国家管理 a12a 的全局领域句类表示式设计如下：

SCD=SCD(a12a)=SCD(a12a)=ReBD01-42R411-e61\*211J+~R711-e63J

其中各语义块的对应如下：

ReB=a50\3ga12a(国家管理法规); DA=pea119(pj2/pj2-0)//pea119-0(国家管理者)

RB2B1=pj2-0//pea119-0(管理机构);RB2C1=a//d(管理对象从事的行业或活动);

RB2B2=pj2-0//pea119-0(管理机构);RB2C2=a//d(管理对象从事的行业或活动)

管理在任何组织机构中都需要，国家同样需要管理，国家管理体现治国所需的一般性管理知识。管理需从纵向和横向进行考虑，横向机构具有地域的分布性和实施的执行性，纵向机构具有地域的集中性和实施的决策性，这一特性在政治管理中表现得最为明显。政治管理的最高横向侧面对应着地方政府，最高纵向侧面对应着中央各部委。一个侧面体现由 DA 对不同下一级政权机构 RB2B 的管理，管理过程中同样体现出局部考虑的谋略智慧 D01，同时另一个侧面体现平级之间管理者的相互协调管理的问题。管理的对象都为 pj2-0//pea119-0，管理的内容都是管理对象从事的专业活动 a//d；同时治国一定具有相关的法律法规，要根据不同的分类区别对待，故在该领域常常出现参照辅块 ReB。

## 2 具体示例

以上以专业活动 a 行的概念树治国 a12 的一级延伸概念 a12t=a 为例，说明了该领域的领域句类表示式的设计及其所应该体现的世界性知识，下面通过句群例子来说明全局领域句类表示式是如何具体使用的。

例：“罗斯福新政”的意义|在于，||{它|开创了|资本主义的改革之路}。++为了{完善|资本主义制度}~||“新政”||突破了||亚当·斯密以来自由主义的传统模式，+大胆地借鉴||社会主义的|计划经济和福利制度的|优先之处，+并部分地将其||纳入||制度框架，+形成了||<<不同于|苏联“战时共产主义”|的国家资本主义的新形式>>，+确立了||美国式的“政府干预—福利国家”的改革模式<sup>[3]</sup>。

句类分析结果如下：

SC=P22J#PBC1={XY10\*21J}++RtC&Y10J#RtC={!31XY50\*21J }+D01R511\*21J+!1111R011T0\*32J+Y10S0\*21J#SC=<1242jD00J>+Y10J

这一句群，围绕“罗斯福新政”重要意义展开，体现了开拓性治理这一领域的句类表示式的设计思想，文体为论述体。“新政”一词的 HNC 映射符号为(ra109,r8218)，符号就有很大的领域特点，既可属于政治 a1 下的制度与政策 a10 概念树，表示政策的静态结果，也可属于治国 a12 概念树，表示“开拓性治理”的效应结果，但表达这两个领域所用的核心词是不同的。句类分析结束以后，通过特征语义块表达词语“完善(v900#v50051e511)、突破(v93129)、借鉴((v93119,v9451)+114)、形成(v311,v30ae71)、确立(v9351)”的 HNC 映射符号可知，描述主体都是动态描述，由此可判断本句群的“新政”属于“开拓性治理”的效应结果，排除 a10

概念树静态结果的可能性。领域确定后,计算机即可由领域知识库获取对应的 a129 领域句类表示式,掌握对应领域世界知识。虽然在句类分析的句类代码上并不完全等同 a129 领域句类表示式(见上文),但领域句类表示式中各个部分都在具体例子中得到了展现。领域句类表示式中的因果辅块 RtC 仍旧通过 RtC 来体现;参照辅块 ReC 通过小句(第 2 个小句“大胆地借鉴社会主义的计划经济和福利制度的优良之处”)来体现;时间辅块 Cn-1 并没有体现,而是在文章的最开始部分通过“罗斯福新政”的总体介绍时表达出来了;而特征语义块部分(D01-42Y101-1\*211J)则通过各特征语义块部分得到了体现。通过围绕“新政”而展开的“完善、突破、借鉴、形成、确立”这一系列词语的搭配体现了 a129 这一延伸结构的全局领域句类知识。

对应的领域句类表示式如下:

SCD=SCD(a129)=Cn-1|RtC(ReC)D01-42Y101-1\*211J

对比句类分析结果和领域句类表示式结果发现:领域句类表示式中的因果辅块 RtC(“为了完善资本主义制度”)、Y101(“开创”)、YCB(“美国”)、YCBC(“资本主义制度”)、参照辅块 ReC(“借鉴社会主义的计划经济和福利制度的优良之处”)都出现了,但直接以领域句类表示式中的方式出现的仅有 RtC,而谋略者(DA)和时间辅块(Cn-1)则未在句类分析表示式中出现,需要通过背景知识库来补充相关语义块信息。

在语境单元萃取过程中,可从中获取的信息如下:

SGUD=(8y:|DOM;SIT;BACE;BACA)

其中:8y=841(评价);

DOM=a129(国家开拓性治理);

SIT=SCD(A,B,C); DA(罗斯福)未出现、YCB(美国)、

YCBC(资本主义制度);

(上接第 188 页)

6)将该案例的基本信息(如异常类型、发生时间、地点、影响范围、执行者等)和相关信息(如异常处理方式、处理时间及其执行者等等)迅速提交系统管理员,由管理员根据情况做出方案判断和调整,采用相关措施进行异常处理;

7)上述异常处理完成后,将其相关处理内容(异常发生地点、时间、类型、采取措施等),按前述 ECA 规则构成形式,形成新的 ECA 规则,加入规则库。

(2)当找到一致项时

1)若找到匹配项 ECA<sub>i</sub>,则继续判断 ECA<sub>i</sub>.Co 是否满足;

2)若不满足,则继续判断直至条件满足;

3)若满足,则选择 ECA<sub>i</sub>.Ac.Oper 中的类型(包括忽略 Ignore、重试 Retry、修改实例 ModifyInstant、修改模型 ModifyProc、后向恢复 BackReset 和组合 Comp6 种处理操作,后 4 种操作需要调用 ECA<sub>i</sub>.Ac.Proced 进行具体处理)。

## 5 结论

综上所述,PK-BPM 系统的异常处理机制是:对于可预料的异常,根据 ECA 规则库中已有的处理策略和方式,执行相应的处理操作,如忽略、重试等;而对于那些库中没有与之匹配规则的异常,则按相似度值最小的定则在库中找到与本次异常最相似的案例提交给系统管理员,再根据实际情况对案例进行相应调整后,将其作为本次异常的处理方案。并在处理完成后,将这一过程写成新的 ECA 规则,以备后续使用。

实际中大部分异常都是相似于 ECA 库中的异常而非相同,PK-BPM 系统利用流程中的异常处理知识,对 ECA 法进行扩展,使系统的异常处理能力得到了较大提升,系统的柔

BAC[E//A]:需要更加专业的知识库提取(关于罗斯福新政的背景知识库)

通过以上示例,体现了治国领域的领域句类表示式如何帮助计算机理解,这将对计算机句群处理具有极大帮助。其它领域的句群都可以通过这一方式得到相应的形式化知识库。关于 SGUD,可见文献[2]。

## 3 结束语

人类语言的模糊性和计算机语言的确定性之间的矛盾造成了语言空间的句群无法一一对应到概念空间的语境单元,在句类分析完成以后,必须通过词语 HNC 映射符号之间的聚集性来判断句群的领域,通过领域认定确定语境单元的范围。本文立足于“世界知识可先验赋予计算机掌握”这一论断,通过领域句类表示式的设计及配置,建立语境单元萃取所需的领域句类知识库,使句群处理能得到世界知识支持,并最终使得句群处理得到较大突破。文中仅以概念树治国(a12)的一级延伸概念 a12t=a 的领域句类表示式为例,实际 a 行其它各概念树的领域句类表示式已设计完成,并通过了示例句群的测试,效果良好。而其它领域的领域句类表示式的设计将是我们下一步开展的工作。

## 参考文献

- 1 黄曾阳. HNC(概念层次网络)理论[M]. 北京:清华大学出版社,1998.
- 2 黄曾阳. 语言概念空间的基本定理和数学物理表达式[M]. 北京:海洋出版社,2004.
- 3 刘灼献. 论 20 世纪以来西方主流意识形态的演变及其功能[J]. 河南大学学报,2004,(3).

性也因此而得以改善。

## 参考文献

- 1 Harmon P. IBM's BPM Strategy, Products and Architecture[J]. Business Process Trends, 2004, 2(11).
- 2 Li Weiping, Xue Jinsong, Zhu Yunlong. The Dynamic Modeling Method of Workflow Management System Supporting BPR[J]. China Mechanical Engineering, 2002, 13(15): 1314-1317.
- 3 朱锦泉,苑森森. 基于事务的工作流异常处理模型及方法[J]. 吉林大学学报(工学版),2003,33(3): 46-50.
- 4 孙瑞志,史美林. 工作流异常处理的形式描述[J]. 计算机研究与发展,2003,40(3): 393-397.
- 5 杨东,张申生. 基于“多层日志”的工作流的失败处理[J]. 计算机集成制造系统,2003,9(4): 299-304.
- 6 赖耀东,朱建新. 基于多 Agent 虚拟组织工作流管理系统的异常处理机制[J]. 计算机工程,2003,29(2): 68.
- 7 Luo Z, Sheth A, Kochut K, et al. Exception Handling in Workflow Systems[J]. Applied Intelligence, 2000, 13(2): 125-147.
- 8 Hagen C, Alonso G. Exception Handling in Workflow Management Systems[J]. IEEE Transactions on Software Engineering, 2000, 26(10): 943-958.
- 9 Chiu D K W, Li Qing, Karlapalem K. Web Interface-driven Cooperative Exception Handling in ADOME Workflow Management System[J]. Information Systems, 2001, 26(2): 93-120.
- 10 Klein M, Dellarocas C. A Knowledge-based Approach to Handling Exceptions in Workflow Systems[J]. Computer Supported Cooperative Work, 2000, 9(3): 399-412.