文章编号：1003-0077(2018)04-0066-08

基于概念层次网络的知识表示与本体建模

文 亮，李 娟，刘智颖，晋耀红

（北京师范大学 中文信息处理研究所，北京 100875）

摘 要：知识表示是自然语言理解的重要基础。知识表示不统一、语义信息无法系统化利用是目前存在的亟待解决的问题。要解决这个问题，就要解决语义知识表示的问题。该文基于概念层次网络，描述了词语、句子和篇章层面的语义知识表示方法。基于文中描述的词汇层面的表示方法，构建了一个多语言本体知识库。该知识库的知识表示方法不仅可以为知识表示理论研究提供基础，还可以为自然语言处理相关领域的应用提供资源支持。

关键词：概念层次网络；语义知识表示；本体建模

中图分类号：TP391 **文献标识码**：A

A Method of Knowledge Representation and Ontology Modeling Based on Hierarchical Network of Concepts

WEN Liang, LI Juan, LIU Zhiying, JIN Yaohong

(Institute of Chinese Information Processing, Beijing Normal University, Beijing 100875, China)

Abstract: In the field of natural language processing (NLP), it is an important issue to be addressed at present that knowledge representation is not unified, and semantic information can not be used systematically. This paper presents a multi-dimensional and unified semantic knowledge representation method covering words, sentences and discourses based on the theory of hierarchical betwork of concepts (HNC). With this method, we build a multi-language ontology based knowledge base (KB), which can provide theoretical reference for the semantic processing study of large-scale Chinese texts, and support for the construction of knowledge resources in specific domains.

Key words: hierarchical network of concepts; semantic knowledge representation; ontology modeling

0 引言

在自然语言处理（NLP）领域，知识表示(knowledge representation)的主要目标是把知识数字化、形式化、系统化，便于计算机储存、识别、理解和处理知识。知识表示是自然语言理解的前提和基础，任何语言的理解都要建立在知识表示的基础上。

在人工智能领域，本体(ontology，又称为本体论)是一种“形式化的，对于共享概念体系的明确而又详细的说明”[1]。本体提供的是一种共享词表，也就是特定领域之中那些存在着的对象类型或概念及其属性和相互关系[2]。所以，本体实际上是依据某种类别体系，对实体、概念、事件及其属性和相互关系的形式化表达。

概念层次网络(hierarchical network of concepts, HNC)[3]理论以概念联想脉络为主线，建立了一种模拟大脑语言感知过程的自然语言表述、理解和处理模式，使计算机获得消解歧义、理解自然语言的能力。HNC 通过类别符号、层次符号以及结构符号的组合，构建了自然语言概念空间的符号化表述体系，可以表述词语、句子、句群和篇章层面的语义知识；同时，HNC 以概念基元为基本单位，概念基元的联想脉络模拟了人脑的认知机制，可以实现概念之间的激活、联想、扩展、浓缩和存储功能。

本文基于概念层次网络的知识表示方式，构建了多语言本体词语知识库。具体来说，是以 HNC 概念节点表为纲，对每一个概念进行文字解释，并列

收稿日期：2017-03-03 **定稿日期**：2017-05-16

基金项目：国家高技术研究发展计划(863)(2012AA011104)；国家语委“十二五”科研规划项目(YB125-124)

出概念所对应的多语言词语，目前为中英双语词语捆绑。

1　相关工作

目前的知识表示方式主要有两种方式：(1)以 WordNet[4]、知网(HowNet)[5]等本体知识库为代表的知识表示方式；(2)以 Word Embedding 为代表的词向量的知识表示方式。

WordNet 是一个包含了语义信息的机读词典，它能够支持自动文本分析以及人工智能应用。首先，WordNet 描述了每一个词的基本意义；然后，根据词条的意义，WordNet 将具有相同意义的词条集合为一个 Synset(同义词集合)；其次，WordNet 描述了不同 Synset 之间的语义关系。但是，WordNet 只描述了名词、动词、形容词和副词组成的同义词网络，既不深入到义素分析中的义原(primitive)或概念，也不扩展到超越单词层面的脚本(script)或框架(frame)，其描述的语义信息和关系相对有限，有其不足之处。

知网是一个描述词语(汉语和英语)所代表的概念，揭示概念与概念之间以及概念间各种关系的常识知识库。知网定义了事件、万物、属性、属性值、部件、空间和时间七类最顶层的概念。建立了这七类概念之间的关系。知网通过 800 个“义原”对这些概念进行描述。义原指的是最基本的、不能再分割的表达意义的最小单位。为了描述概念间的关系，知网定义了同义、反义、对义、上下义等语义关系。但知网对概念的定义过于模糊，使用义原解释概念，虽然有利于整合概念之间的关系，但这种描述语言的方式不够形式化和结构化，在计算机处理语言时不能很好地被利用。

词向量的知识表示方式一种是 one-hot representation，另一种是 distributed representation，Tomas Mikolov 等提出的词向量表示工具 Word2Vec[6]很有代表性，它将词语转化为向量，之后，Tomas Mikolov 团队也将其推广到了句子和文档的表示中[7]，将它们转换为一个低维语义空间中的数值向量。其优势在于将自然语言处理过程中的语义鸿沟现象，通过低维空间中向量间数值计算得以一定程度的改善或解决[8]，因此基于深度学习知识表示技术在自然语言处理领域得到了广泛应用。但是，向量表示难以具体描述具体的语义信息，在消解歧义方面还面临着巨大的挑战[9]。

基于概念层次网络的知识表示体系和其他知识表示方式相比，该体系以语言理解基因为核心，综合语义和语境信息，描述跨越词汇、句子、句群篇章多个层面的、统一的语义知识表示方法，解决语义信息系统化问题。这种表示体系可以解决面向海量文本处理时，知识表示不统一、语义信息无法系统化利用的问题。不仅可以为大规模中文语义处理核心关键技术和应用系统研究提供理论基础，建设的知识库也可以为面向领域的知识资源建设提供支持。

同时，HNC 多语言本体表示方式以数字化、基元化的概念表示为基础，给出概念之间的关联性、句子的表述模式、句群和篇章的表述框架，以及概念在句子、句群和篇章中的语义、语用信息。语言理解基因不仅可以激活词汇之间的语义计算，也可以激活句子层面的关联计算，同时可以激活句群和篇章层面的语境计算，把大规模文本内容转换为动态记忆。将知识推理蕴含于符号表示之中，与其他工作相比具有独特性与优势。

2　基于概念层次网络的词汇层面知识表示

2.1　概念层次网络

概念层次网络(hierarchical network of concepts，HNC)是模拟大脑对语言感知的过程建立起的表示概念联想脉络的语义网络[10]。这个理论框架是以语义表达为基础的，它对语义的表达是概念化、层次化、网络化的，所以称它为概念层次网络理论[11]。

HNC 理论认为概念无限而概念基元有限、语句无限而句类有限、语境无限而语境单元(理解基因)有限、显记忆无限而隐记忆有限，所以 HNC 将语言概念空间分为概念基元空间、句类空间、语境单元空间、语境框架空间四个层级。HNC 对这四层级的结构体设计了相应的符号体系，建立了语言概念空间体系(包括语义概念基元体系和语句基元体系)，通过作用效应链，建立起层次性、网络性的概念表述模式，从而使计算机能够理解词语、句子、句群及篇章的语义。

2.2　词汇层面的表示模式

2.2.1　语言概念空间符号体系

词汇层面的表示模式主要通过概念节点来表

示，对应于概念基元表示式，即概念基元符号体系。这种表示模式具有语义完备性，能够与自然语言的词语建立起语义映射关系。同时，它高度形式化，每一个符号基元（每个字母或数字）都具有确定的意义，可充当概念联想的激活因子。

HNC把概念分为抽象概念和具体概念。具体概念是指必须确定“所指对象”的概念，基本物概念和挂靠概念属于具体概念，如光和房子；抽象概念是指不必确定“所指对象”的概念，除了基本物概念和挂靠概念的都属于抽象概念。

抽象概念的第一子类即作用效应链，HNC命名为主体基元概念，黄曾阳先生认为“所谓一个事物的知识表示，归根结底就是对作用、过程、转移、效应、关系和状态这六个侧面的表述”[12]，这六个节点是自然语言对万事万物进行描述的六个基本角度，也是一切事物发生、发展和消亡的六个基本环节。在这六个一级节点之下，衍生出许多子节点，共同描述每个概念的不同方面。

抽象概念的第二子类为“扩展基元概念”，主要描述人类活动的方方面面，包括生理本能活动、心理活动及精神状态、思维活动、社会活动等一级节点及其衍生的子节点。HNC理论用五元组特性表示抽象概念的特性。现代汉语将词分为动词、名词、形容词、副词等词性，HNC理论用五元组来描述同一概念的不同侧面，分别代表概念的动态（v）、静态（g）、值（z）、属性（u）和效应（r），具体如表1所示。

表1 抽象概念的五元组特性

HNC符号	HNC说明	词类命名对应	举例
v	概念的作用（动态）描述	动词	思考 v80
g	概念的作用（静态）描述	名词	思维 g80
u	概念的属性描述	形容词或副词	弱 u00c21/强 u00c22
z	概念的值描述	量词	力度 z00
r	概念的效应描述	名词	想法 r80

具体概念中，基本物概念节点主要包括热、光、声、电磁、微观基本物、宏观基本物和生命体这些一级节点及其衍生子节点，但基本物只是具体物的一小部分，挂靠概念也用来描述具体物。挂靠指把一个概念的层次符号与相关概念的层次符号拼接在一起。例如，表示“教师”这个具体物，首先p代表人，其次基元概念的a行是专业活动，所以就将p（人）和a71（a代表专业活动，a7代表教育，a71代表教）的层次符号拼接在一起，pa71就代表“教师”。

HNC使用英语字母、数字、组合结构符作为概念或概念基元的表示符号。描述抽象概念的字母主要有j（表示基本概念）、l（语法逻辑概念）、f（语习逻辑概念）、s（综合逻辑概念），抽象概念具有五元组特性（字母表示如表1所示）；描述具体概念的字母主要有p（人）、w（物），这些字母表示的符号称为类别符号。数字0～14表示概念的层次性内涵，称为层次符号。HNC定义了12种概念组合符，即：作用（#）、效应（$）、对象（&）、内容（|）、偏正（/）、主谓（‖）、展开（+）、并（,）、选（;）、一般逻辑组合（lyy）、非（!）、反（^），这些字母用来表示符合概念的组合结构。

HNC对自然语言概念的符号化表述可以一般化为：

$$\sum\{\text{类别符号串}\}\{\text{层次符号串}\}\{\text{组合结构符号}\}\{\text{类别符号串}\}\{\text{层次符号串}\}$$

类别符号串和层次符号串构成一个概念基元的表达式，组合结构符号可以将两个或多个概念基元组合成新的概念。

例如：“思考”的表达式v80，v代表类别符号，表示这个概念是动态的作用，80代表层次符号，8表示思维活动，80是8的子节点，表示一般思维活动。“阻碍”的表达式为v376 # v362，v376表示阻碍，v362表示抑制，前者是作用，后者是该作用产生的效应，#表示作用产生了后面的效应，组合起来就表示阻碍这个概念。

基于HNC的词语表示在计算语义距离时非常方便，如“国家”表示为pj2，“亚洲国家”表示为pj2 * 1，“中国”表示为pj2 * 16，从它们的HNC表达式可以看出“国家”和“中国”之间是有关联关系的。其中，p表示人，pj表示人化的基本概念，数字表示概念的层次性。

人工生成HNC符号的效率和成本很低，在应用过程中，也产生了HNC符号与词汇的映射工具[13]，这一自动化映射工具大大减轻了词汇与HNC符号的转换成本，为后续的词汇理解、句子理解、句群和篇章理解奠定了基础。

2.2.2　语言理解基因

语言概念空间符号体系的数字化表示是语言理解基因的基础结构，语言理解基因主要靠词语直接激活，有了词语层面的激活才有语句和篇章层面上层建筑的实现。

语言理解基因的总体设计思路可以用如下语言表述：

理解基因::=范畴表示＋结构与功能的各级综合表示（::=表示等价于）

范畴描述层次性；结构与功能描述网络性（关联性）。下文以多语言本体知识库构建为例实现基于语言概念空间符号体系的本体构建。

2.3　多语言本体知识库构建

2.3.1　多语言本体知识库构建的具体标准

2.3.1.1　概念节点的选择

HNC 语义网络中任何一个节点都代表一个概念，同时也都是概念的基元。虽然在现实生活中概念是无限的，但作为概念的“元素”的基元是有限的，这些概念基元可以组合成无穷无尽的概念，从而描述自然语言的所有概念。

HNC 理论认为大脑自然语言理解基因的直接主体构成大约是 15 000 个的概念基元，这有限的 15 000 个概念基元基本可以描述无限的概念。这项理解基因的探索属于大脑研究的战略性课题，目前 HNC 词语知识库针对性地选取了全部的 5 000 个高层概念节点对它们进行描述，这 5 000 个高层概念节点囊括了约 10 万条词语。

2.3.1.2　标注规范

多语言本体知识库以 HNC 概念节点表为纲，对每一个概念进行符号化表示和详细描述，囊括概念涉及的各个侧面的词语，并且通过概念间的关联表示出概念与概念之间的关系。标注主要从对单个概念节点的具体描述、概念与概念间的关联两方面展开。

1）概念节点的描述

HNC 将概念节点映射为由字母、数字、一些代表组合结构符号组成的 HNC 表达式。表达式的每一个符号都具有确定的意义，充当概念联想的激活因子。如 2.2.1 节所述，HNC 把概念区分为具体概念和抽象概念，抽象概念节点具有五元组特性中的全部或部分属性，每个词语从不同侧面描述这个概念节点的多元性表现。具体概念（除基本物概念外）则通过挂靠的方式来表示。

知识库中描述的概念节点的信息[14]应包括：①该节点的中英文命名，②概念节点之间的层次关系（上位概念、下位概念和同位概念），③该节点所捆绑的词语（动态词语、静态词语、属性词语、值词语、效应词语），④概念之间的关联性。

2）概念关联性

词语知识库中，概念之间具有关联性，概念关联式是语言理解基因的主体信息渠道。关联主要通过节点的定义和结构符号的运用规定节点之间的关系，具体包含以下几类：

（1）概念间的层次性

概念节点之间具有高层、中层和底层之分，高层节点表达概念的层次性，中层节点表达概念的对偶、对比和包含特性，底层概念表达概念的网络性。HNC 语义网络中高层层数是确定的，如 j 类：基本概念，其高层节点的层数是两层，表示为 j0，j1，j2，…，j8。中层节点的例子在自然语言中非常常见，如“强 u00c21”与“弱 u00c22”是对比关系，“对 jgu841”与“错 jgu842”是对偶关系，“年 wj10”“月 wj10-0”“日 wj10-00”之间是包含关系。层次性判断可简化为概念表达式的数字串是否相同，因而语义距离计算的部分问题就可使用逐层比较数字串的方法来解决。

（2）概念间的网络性

概念的网络性分为两种形式：交式关联，链式关联。

① 交式关联指的是两个概念有交叉，即同一概念本体从不同观察角度看到的不同映象。如“死亡”这个概念，从过程看，它是“代谢”的“谢 14e62”；从“效应”看，它是“消失 312”；从状态看，它是“减少 50041e42”，所以过程节点 14e62、效应节点 312 和状态节点 50041e42 是交式关联的。

② 链式关联是作用效应链各环节的因果性表现。例如，“效应的扩展与缩小 vg34”链式关联于“量与范围 j4”。

（3）概念关联符号定义的关联性

上述几种关联类型主要通过概念节点本身的表征符号来揭示概念之间的关联性。除此之外，HNC 理论还定义了常见的 10 种逻辑关联类型，并设计了特定的关联符号将概念关联起来，用于描述概念之间的内容逻辑关系。

比如，关联符号“＝%”表示一个概念包含另一个概念。具体的关联符号及其含义如表 2 所示。

表 2 概念关联式的 10 个特定内容逻辑符号

符号	汉语说明	表示含义
=	强交式关联于	表示两概念间存在足够大的交集
=＞	强源式关联于	表示两概念之间的源流(因果)关系
＜=	强流式关联于	表示两概念之间的源流(因果)关系
≡	强关联于	表示两概念间存在着最大的交集和最强的因果关系
:=	对应于	表示两概念具有对应关系
==	虚设	针对延伸概念和概念树
=:	等同于	表示两概念语言意义的等同
%=	属于	表示一概念属于另一概念
=%	包含	表示一概念包含另一概念
::=	定义于	表示一概念等价于另一概念

2.3.1.3 标注一致性

针对选取的 5 000 个高层概念节点,我们希望尽可能地根据概念找到概念所描述的所有词语,将描述它的不同侧面的词语穷尽性地填写在知识库中。知识库的每个概念由两个不同的填写者进行填写,经过对比,对填写者不确定或两位填写者标注不一致之处进行讨论,经过讨论决定最终标注结果。

根据以上的标注规范,我们对选取的 5 000 个高层概念节点进行了描述,具体实例以节点“3a1”即概念“获得”来展示,如表 3 所示。

表 3 概念“获得”的具体描述

属性名	属性值
概念节点	3a1
中文命名	【获得】
英文命名	obtain
上位概念	3a【获得与付出】
下位概念	3a13【不道德的获得】;3a19【需求】;3a1a【索取】
同位概念	3a2【付出】
动态词语	获得; 博得; 捕获; 得到; 取得; 赢得; 攫取 obtain; receive; gain; achieve; win; get; procure; attain; acquire
静态词语	obtaining; procurement; acquisition
值词语	获得性
效应词语	得分; 薪水; 收入; 税收; 成果 score; payment; achievement; tax; outcome
属性词语	available; obtainable; handy
挂靠类型	Null
具体概念	Null
关联式	::=
关联节点	(201∪3818)$461

概念节点“3a1”的中文命名为【获得】,英文命名为“obtain”。

概念【获得】的形式化表示符号为“3a1”,其上位概念为“3a【获得与付出】”,下位概念为“3a13【不道德的获得】;3a19【需求】;3a1a【索取】”,同位概念为“3a2【付出】”。

概念关联的五元组中动态词语为“获得；博得；捕获；得到；取得；赢得；攫取 obtain；receive；gain；achieve；win；get；procure；attain；acquire”，静态词语有“obtaining；procurement；acquisition”，值词语为“获得性”，效应词语为“得分；薪水；收入；税收；成果 score；payment；achievement；tax；outcome”，属性词语为“available；obtainable；handy”。

挂靠类型和具体概念这两处为空值。

关联式为“::=”表示节点【3a1】等价于关联节点【(201∪3818)$461】。

通过表 3 中各项信息的描述，“获得”这一概念就以概念层次网络的表示方式被描述出来了。

2.3.2　知识库中概念的更新

HNC 理论认为概念无限而概念基元有限，现有的 HNC 概念符号能够表示任何概念，而具体概念向抽象概念挂靠，新出现的具体概念可以通过向抽象概念挂靠实现。

目前，本体广泛应用的一个瓶颈在于本体构建的自动化程度不高，多数本体还依赖于手工构建。如何提高本体构建的自动化程度，减少本体构建的成本，提高本体的共享程度，是目前亟待解决的问题。我们所构建的多语言本体知识库是一个动态更新的系统，填写者可以按要求填写概念知识，管理员经过审核后可以确认删除或修改填写的概念节点。我们希望不断有新的填写者加入本体知识库的构建中，采用众包的方式，不断扩展、完善知识库，使之成为能被调用的活知识。填写界面如图 1 所示。

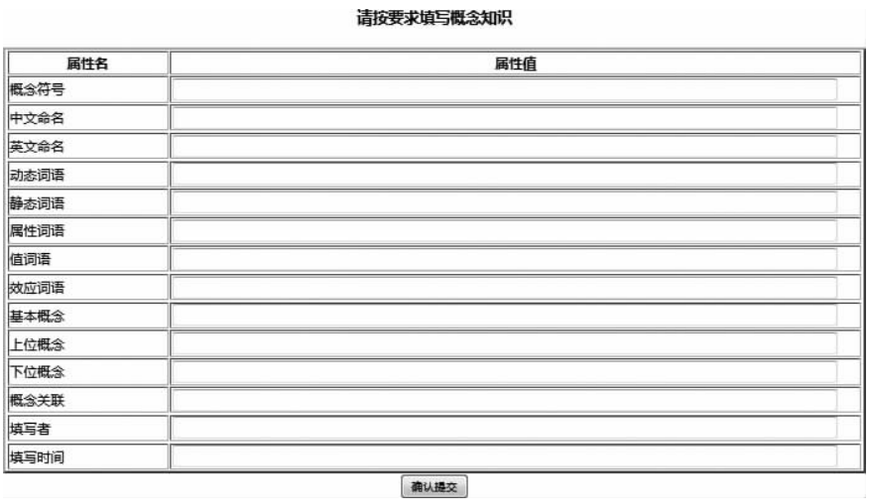

图 1　多语言本体知识库中概念知识填写细目

填写者可以填写概念符号的属性值，包括中英文命名，此概念捆绑的动态词语、静态词语、属性词语、值词语、效应词语（填写的词语需有中英文对照），基本概念、上下位概念和概念关联。

2.3.3　多语言本体知识库的应用

多语言本体知识库目前已应用到机器翻译的实际任务中，可解决汉英概念之间的映射问题，这种映射不单单只是词语之间的映射，而是两种自然语言之间的转换，这种自然转换可以提高机器翻译系统的译准率。同时数字化、符号化的词语表示方式对于语义距离的计算很有优势，在选择候选词时，系统能够根据 HNC 编码优先判定常用搭配语块。

3　基于概念层次网络的句子层面知识表示

句子层面的知识表示模式是指用句类表示式描述句子的语义结构特征，HNC 用句类（sentence category，简称 SC）表示式来表征无限的语句。HNC 定义的句类指的是句子的语义类型，而不是指陈述句、疑问句、祈使句和感叹句之分[15]。句类体系主要由广义作用句和广义效应句组成，前者包括作用句、转移句、关系句和一般判断句四个类型，后者包括过程句、效应句、状态句和基础判断句四个类型[16]。这八大类型细分为 57 种基本句类，57 种基本句类理论上可以衍生出 3 192 组混合句类。以 57 种基本句类为基元，通过句类的混合和复合就可

以实现对自然语言语句的语义结构描述。句类命名和句类符号对应关系如表4所示。

表4 句类命名和句类符号对应关系

句类命名	作用句	过程句	转移句	效应句	关系句	状态句	判断句
句类符号	X	P	T	Y	R	S	D

句类表示式由语块构成,语块是语句的下一级语义构成单位。HNC定义语块是句类的函数,即句类决定句类表示式中含有哪些语块的表示式。语块存在主块和辅块两种基本类型,语块和主块用同一个字母K表示,辅块用字母fK表示。主块四要素为:特征要素(E)、作用者(A)、对象(B)和内容(C),辅块七要素为:手段(Ms)、工具(In)、途径(Wy)、比照(Re)、条件(Cn)、起因(Pr)、目的(Rt)。

HNC句类一般表示式如下:

SC=JK1+{EK+JKm}(m=2~4)

SCR= SC+fKm

举例如下:

例1 李四||拒绝了||领导的要求。

X21J=X2A+X2+XBC

主动反应句=反应者+反应行为+反应引发者及其表现

例子中,X21是句类代码,X表示作用句,等号右边是这个句子的句类表示式。其中,X2A表示反应者,X2表示反应行为,XBC表示反应引发者及其表现。

主动反应句属于广义作用句,还可以有不同的格式代码,例子可以变为"李四把领导的要求拒绝了(! 11X21J=X2A+XBC+X2)"、"领导的要求被李四拒绝了(! 12X22J=XBC+X2A+X2)"。

通过字母符号及句类衍生,HNC句类表示式可以实现对有限的句类的表示,从而解决无限的语句形式化问题。

4 基于概念层次网络的篇章层面知识表示

在HNC表示体系下,我们把信息抽象成三个侧面:领域、情景、背景,三个侧面构成语境三要素[17]。(在这里,我们把句群、段落、篇章称为信息的载体。)对句群、段落、篇章的表示就是对不同颗粒度大小的语境的描述。通过对表征信息的三个不同侧面(领域、情景、背景)的描述,我们可以形式化地表示出语境。

在HNC语境框架理论中,领域描述事件的所属类型,可以看成是对事件范畴的静态描述。情景用来描述事件的作用效应链的具体表现。各参与者以及他们之间的语义关系、事件的内容通常由情景描述指定。背景则用来描述事件发生的条件、叙述者和论述者的背景、叙述者和论述者的特定视野等。情景和事件背景可以理解为是领域的函数。

HNC理论认为,任何语段、篇章等构成的语境都是由若干个有限的基本构件组合而成。我们把这些基本构件称为语境单元。语境单元由领域DOM、情景SIT和背景BAC三要素构成,而背景BAC又分为事件背景BACE和述者背景BACA。语境框架被用来抽象表示语境各要素的构成方式。语境各要素的构成方式可以形式化地表示如下[18]:

SGUN=(DOM;SIT;BACE;BACA)

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

SIT=SCD(A,B,C)

其中,SGUN—语境单元,分为叙述(Narrate)型、论述(Discuss)型;DOM—领域;SIT—情景;BAC—背景;BAC[E//A]——事件//述者背景;SGUD—语境框架;SCD—领域句类。

语境描述的基础来源于对上下文词语的HNC概念符号的解析。在HNC中,概念基元体系网络中的扩展基元概念专门用来描述人类活动。人类不同的领域活动由不同的符号表示。HNC定义了11大类的领域,每一大类都可以有不同的子类,不同的子类也可以进行组合。语境三要素中的领域信息可以通过解析相关词语的HNC语义符号得到。在确定领域信息后,根据不同领域所蕴含的世界知识,通过进行HNC特有的语义句类分析就可以形成对领域句类SCD的判定。此后,再利用人类专家设计完成的领域句类知识为指导,我们就可以确定语境的情景SIT描述。另外,在领域句类知识的指导下,通过分析辅语义块或某些HNC语义符号,我们就可以用HNC符号形式化地描述出背景BAC。语境的三要素(领域、情景、背景)确定之后,语境的表示也就自然出来了。

5　结语

本文构建的多语言本体词汇知识库可以作为自然语言理解系统的基础资源，应用于信息检索、自动问答、机器翻译等领域。相较于 WordNet 和 HowNet，HNC 词汇知识库是完全符号化、数字化的，具有形式化、层次化、网络化的特点，在具体应用及任务中更加便于计算机分析和处理自然语言。

基于概念层次网络的知识表示方法能更好地解决自然语言歧义性这一难题，本文描述了概念层次网络多个层次（词汇、句子、句群、篇章）的语义知识表示方式，限于篇幅和实际描述的浩大工程，本文对词汇层面的知识表示方式及其本体实现做了具体描述，对句子和句群及篇章层面只介绍了基本的表示模式，对于其具体实现及应用将另行撰文阐述。

参考文献

[1] Gruber T R. A translation approach to portable ontology specifications[J]. Knowledge Acquisition, 1993, 5(2): 199-220.

[2] Fensel D. Ontologies[M]. Berlin and Heidelberg: Springer, 2001: 11-18.

[3] Liu Z, Hu R, Jin Y, et al. The multi-language knowledge representation based on hierarchical network of concepts[C]//Proceedings of the 16th Workshop on Chinese Lexical Semantics. Springer International Publishing, 2015: 471-477.

[4] Miller G A. WordNet: A lexical database for English [J]. Communications of the ACM, 1995, 38(11): 39-41.

[5] Dong Z, Dong Q. HowNet Chinese-English conceptual database[R]. Technical Report Online Software Database, ACL, 2000.

[6] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. Proc of ICLR. 2013, arXiv: 1301.3781.

[7] Le Q V, Mikolov T. Distributed representations of sentences and documents[C]//Proceedings of the ICML 2014, 2014: 1188-1196.

[8] 刘康，张元哲，纪国良，等. 基于表示学习的知识库问答研究进展与展望[J]. 自动化学报，2016，(06)：807-818.

[9] 刘知远，孙茂松，林衍凯，等. 知识表示学习研究进展[J]. 计算机研究与发展，2016(02)：247-261.

[10] 黄曾阳. HNC 理论全书[M]. 北京：科学出版社，2015.

[11] 黄曾阳. HNC 理论概要[J]. 中文信息学报，1997，11(04)：12-21.

[12] 黄曾阳. HNC 的发展和未来[C]. HNC 与语言学研究学术研讨会，2001：53-68.

[13] 熊亮，姚娟. HNC 符号与词汇的映射工具的设计[C]. HNC 与语言学研究学术研讨会，2001：368-372.

[14] 李伟. 基于 HNC 理论的本体知识表示研究[D]. 北京：北京师范大学硕士学位论文，2016.

[15] 苗传江. HNC(概念层次网络)理论导论[M]. 北京：清华大学出版社，2005.

[16] 晋耀红. HNC(概念层次网络)语言理解技术及其应用[M]. 北京：科学出版社，2006.

[17] 黄曾阳. 语言概念空间的基本定理和数学物理表示式[M]. 北京：海洋出版社，2004.

[18] 黄曾阳. 语境表示式与记忆[J]. 云南师范大学学报(哲学社会科学版)，2010，(04)：7-14.

文亮(1990—)，硕士，主要研究领域为中文信息处理。

E-mail: wenliang@mail.bnu.edu.cn

刘智颖(1975—)，博士，讲师，主要研究领域为中文信息信息、语言资源建设。

E-mail: liuzhy@bnu.edu.cn

李娟(1990—)，硕士，主要研究领域为中文信息处理。

E-mail: lijuan@mail.bnu.edu.cn