

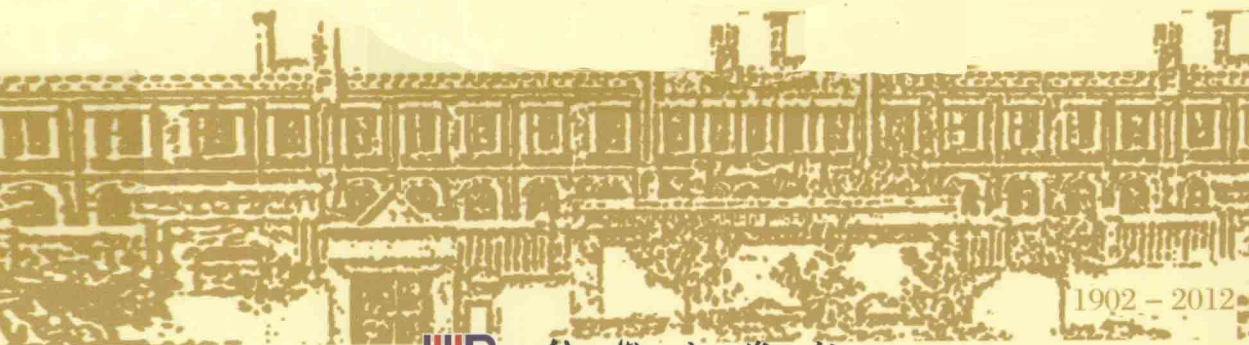


山西大学建校 110 周年学术文库

汉语框架网络本体研究

A Study of Chinese FrameNet Ontology

贾君枝 郜杨芳 刘艳玲 郭丹丹 著



1902 - 2012



科学出版社

山西大学建校 110 周年学术文库

汉语框架网络本体研究

A Study of Chinese FrameNet Ontology

贾君枝 邵杨芳 刘艳玲 郭丹丹 著

科学出版社

北 京

内 容 简 介

本书主要探讨了汉语框架网络本体的构建方法及理论,并对本体在语义 Web 的应用做了相应研究,希望借此研究深化本体与语义 Web 在国内的应用发展。全书共分为六章,具体内容包括语义 Web 与本体,目前常用的词汇本体 WordNet、VerbNet、HowNet、FrameNet,汉语框架网络本体的构建,本体的集成研究,基于汉语框架网络本体的语义检索研究,本体的评估等。本书注重理论与实践研究的统一,提供了丰富的参考文献,以方便读者准确地把握、理解相关专业知识。

本书可作为高等院校信息管理、图书馆学、情报学专业的研究生教材或教学参考书,也可供各类与信息资源处理相关的机构或部门的工作人员学习参考使用。

图书在版编目(CIP)数据

汉语框架网络本体研究/贾君枝等著. —北京:科学出版社, 2012. 5

(山西大学建校 110 周年学术文库)

ISBN 978-7-03-034004-7

I. ①汉… II. ①贾… III. ①汉语-知识库系统-研究 IV. ①TP311.13

中国版本图书馆 CIP 数据核字 (2012) 第 065700 号

责任编辑:郭勇斌 邹 聪 程 凤/责任校对:张怡君

责任印制:赵德静/封面设计:李恒东 无极书装

编辑部电话:010-64035853

E-mail: houjunlin@mail. sciencep. com

科 学 出 版 社 出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

中国科学院印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2012 年 5 月第 一 版 开本: B5 (720×1000)

2012 年 5 月第一次印刷 印张: 14 1/4

字数: 280 000

定价: 48.00 元

(如有印装质量问题, 我社负责调换)

总 序

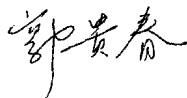
2012年5月8日，山西大学将迎来110年校庆。为了隆重纪念母校110年华诞，系统展现近年来山西大学创造的优秀学术成果，我们决定出版这套《山西大学建校110周年学术文库》。

山西大学诞生于“三千年未有之变局”的晚清时代，在“西学东渐，革故鼎新”中应运而生，开创了近代山西乃至中国高等教育的先河。百年沧桑，历史巨变，山西大学始终与时代同呼吸，与祖国共命运，进行了可歌可泣的学术实践，创造了令人瞩目的办学业绩。百年校庆以来，学校顺应高等教育发展潮流，以科学的发展理念引领改革创新，实现了新的跨越和腾飞，逐步成长为一所学科门类齐全、科研实力雄厚的具有地方示范作用的研究型大学，谱写了兴学育人的崭新篇章，赢得社会各界的广泛赞誉。

大学因学术而兴，因文化而繁荣。山西大学素有“中西会通”的文化传统，始终流淌着“求真至善”的学术血脉。不论是草创之初的中西两斋，还是新时期的多学科并行交融，无不展现着山大人特有的文化风格和学术气派。今天，我们出版这套丛书，正是传承山大百年文脉，弘扬不朽学术精神的身体力行之举。

《山西大学建校110周年学术文库》的编撰由科技处、社科处组织，将我校近10年来的优秀科研成果辑以成书，予以出版。我们相信，《山西大学建校110周年学术文库》对于继承与发扬山西大学学术精神，对于深化相关学科领域的研究，对于促进山西高校的学术繁荣，必将起到积极的推动作用。

谨以此丛书献给历经岁月沧桑、培育桃李芬芳的山大母校，祝愿母校在新的征程中继往开来，永续鸿猷。



二〇一一年十一月十日

前 言

2004年年底,笔者与山西大学计算机学院刘开瑛教授相识,被他孜孜不倦的科研精神感染,参与到其研究团队中,开始真正与 FrameNet (框架网络) 结缘,并被框架语义思想吸引。随着了解的逐步深入,笔者迫切地觉得有必要设计 FrameNet 的汉语版本,因此和刘开瑛教授,以及上海师范大学,山西大学计算机学院、数学学院、管理学院等许多单位的同事一起,参与到汉语框架网络本体的构建过程中。从最初关于 FrameNet 的体验感受,到最终着手建设汉语框架网络知识库,经历了一年多时间。当时语义 Web 及本体论思想已基本成熟,出现了许多著名的本体及本体构建理论;随着 FrameNet 版本的更新,FrameNet 已开始逐步向本体靠近,研究者着手丰富语义关系的研究,并用 OWL (Web Ontology Language, 网络本体语言) 描述框架库及例句库,建立与 SUMO (建议顶层共用知识本体) 的映射。这些信息告诉我们,FrameNet 不是单纯的计算机词典,而是一个计算机形式化描述的语义 Web 词典。它将若干个词元以框架为单元组织成概念,通过丰富框架关系建立概念间的关系,借助框架元素表示概念的属性。FrameNet 涉及医疗、通信、法律、认知等许多领域,考虑到如果完全作 FrameNet 汉化,需要有各类专业人士参与,尤其是例句标注方面,需要投入大量的人力与物力。由于各方面的条件有限,我们决定选择特定领域,并致力于其应用实践研究,在该领域专家的支持下,着手汉语框架网络知识库的构建。山西大学管理学院选择法律领域,并邀请太原市检察院刘建刚、姜忠市检察官,山西大学管理学院孙丽娜老师,负责法律框架的描述;山西大学刘开瑛教授带领的团队所设计的分词软件、框架语义标注规则、半自动语义标注系统为我们开展工作提供了极大的便利;笔者带领的团队应用本体构建理论,着手构建汉语框架网络本体体系,研究语义关系、本体语义类型,提出构建方法,采用 OWL 手工描述,研究数据库到 OWL 文档的自动生成,并考虑汉语框架网络本体与其他本体的集成等问题,研究将框架网络资源作为本体库运用到语义检索实践中,以突破关键词检索的局限性,实现概念层面的检索。这些研究成果的取得得益于及时地追踪 FrameNet 的研究成果和研究者不懈的努力,旨在借此研究深化本体与语义 Web 在国内的应用发展。

本书共分为六章,第一章语义 Web 与本体,描述了语义 Web 的发展历程、特点,以及本体的概念、类型、构建原则及方法,对语义 Web 表示语言——XML (Extensible Markup Language, 可扩展标记语言)、RDF (Resource Description

Framework, 资源描述框架)、OWL 的特点及语法作了介绍。第二章目前常用的词汇本体, 为更好地深入了解 FrameNet, 我们对与其相似的其他词汇本体 WordNet、VerbNet 作对比研究, 从理论基础、组织结构、语义关系、应用范围层面对这些词汇本体进行具体分析, 以明确其各自的理论依据、特点及应用领域。第三章汉语框架网络本体的构建, 提出了针对该本体的具体构建方法, 并依此方法指导本体建设实践, 同时, 为实现对本体进行管理与维护、对语料进行管理及分析, 我们设计开发了本体管理系统、语料库管理系统。第四章本体的集成研究, 探讨汉语框架网络本体、WordNet、VerbNet、SUMO、同义词词典集成问题, 旨在利用其他本体的研究成果, 通过映射实现汉语框架网络本体的词汇、语义类型的扩展, 实现本体之间的重用与共享。第五章基于汉语框架网络本体的语义检索研究, 研究基于汉语框架网络本体资源实现对资源库的框架语义标注及对用户检索提问的语义解析和语义匹配问题, 在检索实践中运用汉语框架网络本体提高对资源的检索效率。第六章本体的评估, 论述了当前本体评估方法及知名的评估工具, 制定了一套本体构建的评估指标体系, 并利用所构建的评估工具对该体系进行了有效性测试。

本书是国家社会科学基金项目“汉语框架网络知识本体构建研究”(06CTQ004)的研究成果, 是研究团队几年来辛勤劳动的结晶。山西医科大学的邵杨芳老师承担了语义检索研究任务, 山西中医学院刘艳玲助理馆员承担了本体评估任务, 太原科技大学郭丹丹助理馆员承担了本体构建的任务, 参与本书编写的还有毛海飞、董刚、董文清、韩笑、卫荣娟、罗林强、赵文娟、牛亚楠、王永芳, 在此对所有团队成员表示感谢。同时, 本书也得到了胡昌平教授、刘开瑛教授的关心与支持。感谢山西大学的领导与同事的关心与帮助, 感谢默默支持笔者的家人, 笔者将拥有这些爱继续前行。

由于笔者水平有限, 书中疏漏在所难免, 敬请读者批评指正。

贾君枝
2011年10月

目 录

i	► 总序	
iii	► 前言	
✓	► 第一章 语义 Web 与本体	
	第一节 语义 Web	1
	一、Web 的发展状况	1
	二、语义 Web 及其体系结构	2
	第二节 本体	4
	一、本体的概念	4
	二、本体的构成	4
	三、本体的类型	5
	四、本体构建	5
	第三节 XML	7
	一、XML 的特点	7
	二、XML 的结构	8
	三、XML 名称空间	9
	四、文档类型定义 DTD	9
	五、XML Schema	12
	六、XSL	14
	第四节 RDF 语言	17
	一、RDF 模型	17
	二、RDF 语法	20
	三、RDF Schema	23
	第五节 OWL	24
	一、OWL 结构	24

	二、OWL 语法	25
33	► 第二章 目前常用的词汇本体	
	第一节 WordNet 本体	33
	一、词汇矩阵模型	34
	二、语义关系	34
	三、词类	36
	第二节 VerbNet 本体	38
	一、VerbNet 描述结构	39
	二、VerbNet 语义关系	41
	第三节 HowNet 本体	42
	一、义原	42
	二、关系	42
	第四节 FrameNet 本体	45
	一、FrameNet 构成	45
	二、FrameNet 框架关系	50
	三、框架元素类型	57
	四、语义类型	63
	五、框架关系中语义类型映射对应	68
75	► 第三章 汉语框架网络本体的构建	
	第一节 汉语框架网络本体构建方法	75
	一、确定本体的应用范围与目的	75
	二、识别领域内部概念及其属性	76
	三、识别概念之间的关系	78
	四、建立领域本体与其他本体的映射	78
	五、编码化处理	78
	六、基于本体应用的评估	78
	第二节 基于法律领域的汉语框架网络本体构建	79
	一、确定法律领域的核心概念集	79

二、构建框架网络本体概念关系	80
三、法律领域的汉语框架网络本体模型	82
四、Protégé 手工描述	82
第三节 汉语框架网络本体管理系统	89
一、汉语框架本体库的存储	89
二、汉语框架网络本体管理系统	94
第四节 法律框架语料库管理系统的设计与实现	96
一、法律框架语料库管理系统的构建原则	96
二、法律框架语料库系统设计	97
第五节 汉语框架网络数据库到 OWL 本体的自动转换	103
一、利用 Jena 构建本体的步骤	103
二、汉语框架网络本体模板构建	104
三、汉语框架网络本体模板的自动生成	109
四、汉语框架网络本体的自动转换实现	114

126 ► 第四章 本体的集成研究

第一节 汉语框架网络本体、VerbNet、WordNet 集成研究	126
一、汉语框架网络本体、VerbNet、WordNet 特点分析	126
二、汉语框架网络本体、VerbNet、WordNet 集成方法	127
第二节 汉语框架网络本体中词汇的扩展	133
一、词元库对汉语框架网络本体应用的影响	133
二、基于同义词表的汉语框架网络本体的词汇扩展	135
三、基于 WordNet 的汉语框架网络本体的词汇扩展	139

第三节	汉语框架网络本体与 SUMO 建立映射研究	144
一、	汉语框架网络本体中的框架、词元与 SUMO 建立映射	145
二、	汉语框架网络本体中的语义类型与 SUMO 建立映射	147
三、	框架元素与 SUMO 建立映射	150
第四节	语义类型的自动确定	152
一、	汉语框架网络本体中框架元素的特点	153
二、	汉语框架网络本体中的语义类型与 WordNet 的体系结构	153
三、	框架元素语义类型自动确定的思路	154
159	► 第五章 基于汉语框架网络本体的语义检索研究	
第一节	本体在信息检索中的应用	159
一、	本体在信息检索中的作用	160
二、	基于本体的语义检索目标	160
第二节	基于本体的语义检索系统设计模型	161
一、	系统的设计思路	161
二、	系统的基本框架	161
第三节	查询子系统中的关键技术	168
一、	问句处理技术	169
二、	句子相似度计算	173
第四节	基于汉语框架网络本体的语义检索实验系统	179
一、	LawOntoSearch 的开发环境	179
二、	LawOntoSearch 的系统功能结构	180
三、	LawOntoSearch 信息检索流程	180

184 ► 第六章 本体的评估

第一节 国外先进的评估工具	184
一、本体评估的意义	184
二、本体评估方法	186
三、国外评估工具的比较分析	186
第二节 本体评估指标体系的构建	192
一、本体评估内容	192
二、本体评估指标的构建	193
第三节 本体评估工具	199
一、本体评估系统基本思路	199
二、本体评估系统基本框架	200
三、本体评估辅助系统——Onto _ evaluate 的特点	200
第四节 基于 Onto _ evaluate 的评估实施	201
一、基于 Onto _ evaluate 的概念评估实施	201
二、基于 Onto _ evaluate 的本体概念间关系 的评估实施	206

209 ► 参考文献

互联网的迅速发展导致信息量呈指数级式增长，但由于 Web（World Wide Web，万维网）页面的无结构性、超链接的自由无序及其搜索引擎智能性不高等缺点，信息查准率和查全率相对较低，人们往往难以获取自己所需的信息。这在较大程度上源于现有搜索引擎基本采用关键字匹配办法，计算机不能理解概念，无法进行语义关联和推理。因此，伯纳斯·李（Tim Berrers-Lee）于 2000 年 10 月 18 日在 XML2000 会议上正式提出语义 Web（Semantic Web）的概念，并随着 2001 年 2 月万维网联盟（World Wide Web Consortium, W3C）组织正式推出语义 Web 活动，网络环境下的语义 Web 进入网络研究发展的主流。在语义 Web 的体系结构中，本体论（Ontology）具有核心地位，语义 Web 的研究与本体论的相互促进已成为共识。本体论通过对概念的严格定义和概念之间的关系来确定概念的精确含义，表示共同认可的、可共享的知识，成为语义 Web 中语义层次上信息共享和交换的基础。基于本体论的方法是基于知识的、语义上的匹配，在查准率和查全率上有更好的保证，对面向 Web 信息的知识检索必将起到关键性的作用。

第一节 语义 Web

一、Web 的发展状况

1989 年 3 月，Web 由欧洲核子研究组织（European Organization of Nuclear Research）的伯纳斯·李第一次提出。其目的在于使科研人员乃至没有专业知识的人能顺利地从网上获取并共享信息。1990 年，伯纳斯·李编写了 HTTP 协议、HTML 语言及第一个 Web 浏览器。1991 年，Web 正式登陆互联网，彻底改变了人类信息交流方式及商业运作方式，人们可以轻松地发布信息，迅速地查找到所需信息。Web 的发展经历了两个阶段：第一阶段是采用静态页面的形式展现网页信息，用户通过修改 HTML 文档内容实现网页的更新；第二阶段是采用 Java script、VB script 技术，通过动态数据库存取，实现同用户的交互，并将存取的数据动态地生成页面，展示给用户，增强了 Web 处理信息的能力，提高了信息搜

索能力，极大促进了电子商务的发展。

随着 Web 的迅速发展，其局限性日益显露，主要体现在以下两个方面。

①HTML语言的局限性。HTML 语言作为目前 Web 的通用语言，其采用预先定义的标签来描述网页信息，简单易用。但这种标记语言的标签集只是对内容的显示格式作了标记，数据的表现格式和数据融合在一起，缺乏针对数据内容的标签。HTML 语言的这种特点决定了 Web 上的信息内容很难被机器理解，从而制约了一些需要对 Web 上的海量数据进行自动化处理应用的开发。Web 上海量的数据要求以一种能够理解数据语义的方式进行交换和管理，当前基于 HTML 的 Web 技术却很难满足这一要求。②搜索引擎的局限性。基于关键词的搜索引擎是目前用户查找信息的主要工具，但 Web 页面的无结构性、超链接的自由无序及计算机不能有效地理解概念、无法进行语义关联和推理，造成信息查准率和查全率相对较低。

由于 Web 在语义表达、信息检索方面存在缺陷，W3C 致力于改进 Web 的工作。1998 年伯纳斯·李首次提出语义 Web 的构想，阐述了语义 Web 的基本思想：利用语义来重新组织、存储和获取信息，使信息变成机器可识别的知识。在 2000 年 12 月 18 日 XML2000 大会上，伯纳斯·李作了题为“The Semantic Web”的重要演讲，正式提出语义 Web 的概念，目的是使 Web 上的信息成为计算机可理解的信息，从而实现机器自动处理信息，同时提出了语义 Web 的体系结构。

二、语义 Web 及其体系结构

语义 Web 不是一个全新的 Web，而是当前 Web 的扩展。伯纳斯·李认为：“语义 Web 是一个网，它包含了文档或文档的一部分，描述了事物间的明显关系，且包含语义信息，以利于机器的自动处理。”（Lee, 1999）W3C 认为语义 Web 是数据网，其提供了允许数据共享及重用的统一框架，将不同来源的数据以统一的格式描述，并建立数据与外部世界的联系，以实现数据的跨组织及跨平台的集成应用。从这些描述中我们可以理解到：语义 Web 在当前 Web 基础上，增加一个语义（知识）层，其通过所定义的语义规范语言及其构建的知识概念结构，使信息被赋予描述良好的含义，成为机器可以识别、交换和处理的语义信息。

语义 Web 是推动未来 Web 发展的核心动力，它包含了许多相关的基础构件。图 1-1 是伯纳斯·李提出的语义 Web 体系结构（Grigoris and Frankvan, 2004），它给出了语义 Web 中的层次关系：基于 XML 和 RDF/RDFS，并在此之上构建本体和逻辑推理规则，以完成基于语义的知识表示和推理，从而使信息能够为计算机所理解 and 处理。

URI 和 Unicode 是语义 Web 的基础，URI 用来唯一标识网络上的资源，Uni-

code 是一个字符编码集，字符集中所有的字符都用两个字节表示，支持世界上所有语言的字符，旨在使语义 Web 成为一个全球信息网络，从根本上解决跨语言字符编码的格式标准问题。

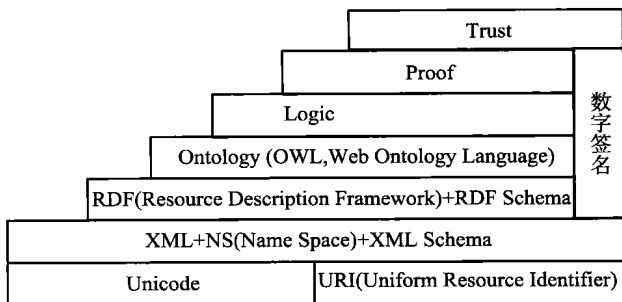


图 1-1 语义 Web 的体系结构

XML+NS+XML Schema 是语义 Web 的语法层。XML 提供了文档结构化的语法，用户可以按照自己的需要定义标记集。为了避免这些标记的同名现象，W3C 引入 Namespaces，即名称空间机制，通过在标记前加上 URI 引用来消解这种冲突。XML Schema 基于 XML 语法，提供多种数据类型，对 XML 标记的结构和使用方法进行了规范，提供了对 XML 文档进行数据校验的机制。

RDF、RDF Schema (RDFS) 属于语义 Web 的数据互操作层。XML 不适于表达数据的语义，数据语义的定义和互操作需要由更高层来完成。因此 RDF 作为基本数据模型，它遵循 XML 语法来描述网络上资源及其关系，提供了一套标准的数据语义描述规范，但它还需要定义描述中使用的词汇。RDFS 提供了一种面向计算机理解的词汇定义，提供了描述类和属性的能力。RDFS 在 RDF 的基础上引入了类、类之间的关系、属性之间的关系，以及属性的定义域与值域等。

Ontology 是语义 Web 的本体层。作为语义 Web 的核心层，本体层在 RDF 和 RDFS 进行基本的类/属性描述的基础之上，更进一步地描述了相关领域的概念集，并提供了明确的形式化语言 OWL，以准确定义相关概念及概念之间的关系。

Logic、Proof、Trust 分别属于逻辑层、证明层和信任层。逻辑层主要解决特定规则的自动推理，定义规则及其描述方法是自动推理的基础，它需要运用功能强大的逻辑语言来实现推理。证明层使用逻辑层定义的推理规则进行逻辑推理，得出某种结论，并负责为应用程序提供一种机制，以决定是否信任给出的论证。信任层保证语义 Web 应该是一个可以信任的网络，它采用数据签名等信息技术来验证信息来源及信息的可靠性，以确保用户信任所见的数据及所作的推理过程。

第二节 本体

在语义 Web 的体系结构中，本体处于核心的地位。本体为语义 Web 提供了相关领域的共同理解，确定了该领域内共同认可的概念的明确定义，通过概念之间的关系描述了概念的语义，是解决语义层次上 Web 信息共享和交换的基础。

一、本体的概念

本体的概念最初起源于哲学领域，可以追溯到古希腊哲学家亚里士多德。他将本体定义为“对世界客观存在物的系统描述，即存在论”，也就是说本体是客观存在的一个系统的解释或说明，它关心的是客观现实的抽象本质。20 世纪 60 年代本体被引入信息领域后，越来越多的计算机信息系统、知识系统的专家学者们开始研究本体，并给出了许多不同的定义。其中最著名并被引用得最为广泛的定义是由 Gruber 于 1993 年提出的“本体是概念化的明确的规范说明”（Grigoris and Frankvan, 2004），1997 年 Borst 进一步将其完善为“本体是共享概念模型的形式化规范说明”（Borst, 1997），Studer 等对上述两个定义进行了深入研究，认为本体是共享概念模型的明确的形式化规范说明（Studer et al., 1998），这也是目前对本体概念的统一看法，其定义包含四层含义。

（1）概念模型（conceptualization）：通过抽象出客观世界中一些现象的相关概念而得到的模型，其含义独立于具体的环境状态。

（2）明确（explicit）：所使用的概念类型及使用这些概念的约束都有明确的定义。

（3）形式化（formal）：采用的是计算机可理解的精确的数学描述，而不是面向自然语言。

（4）共享（share）：本体捕获共同认可的知识，反映的是相关领域中公认的概念集，针对的是团体而非个体。

二、本体的构成

Perez 等认为本体中的知识形式化主要由五部分构成：类（classes）或概念（concepts）、关系（relations）、函数（functions）、公理（axioms）和实例（instances）（Perez and Benjamins, 1999）。

1. 类或概念

通常概念是指任何事务，如任务描述、功能、行为、策略和推理过程。从语

义上讲，它表示的是对象的集合，其定义一般采用框架结构，包括概念的名称，与其他概念之间的关系的集合，以及用自然语言对概念的描述。

2. 关系

关系表示领域中概念之间的交互类型，形式上定义为 n 维笛卡儿积的子集 $R: C_1 \times C_2 \times \cdots \times C_n$ ，如子类关系 (subClassOf)。在语义上关系对应于对象元组的集合。

3. 函数

函数特指一类特殊的关系。该关系的前 $n-1$ 个元素可以唯一决定第 n 个元素。其形式化的定义为 $F: C_1 \times C_2 \times \cdots \times C_{n-1}, \rightarrow C_n$ 。例如，motherOf 就是一个函数，motherOf (x, y) 表示 y 是 x 的母亲。

4. 公理

公理代表领域知识中的永真断言，如声明概念 “Produce” 和 “Produced by” 是互逆的。

5. 实例

实例代表元素。从语义上讲，实例表示的就是对象，对应于现实世界中的具体的个体。

三、本体的类型

按照不同的分类标准，本体的类型提法不一。Guarino 等根据领域的依赖度及其概括程度将本体划分为 4 类 (Guarino, 1998)。

(1) 顶级本体，描述的是最普遍的概念，如空间、时间、物质、对象、事件、行为等，其通常不依赖于某个特定问题或者某个领域，如 Cyc、Mikrokmos、SUMO 本体等属于顶级本体。

(2) 领域本体，描述的是特定领域中的概念，如地球与环境领域本体 SWEET、数字化产权本体 IPRonto。它需要对顶级本体的概念具体化。

(3) 任务本体，描述的是具体任务或行为中的概念及概念之间的关系，如对话本体、销售本体。

(4) 应用本体，描述的是同时依赖于特定领域和具体任务的概念和概念之间的关系。

四、本体构建

本体创建过程非常费时费力，需要一套完善的工程化的系统方法来支持，特

定领域本体还需要专家的参与，以保证所建立的本体符合一定的质量要求，满足用户的应用需求。

1. 本体设计原则

为有效地指导及评估本体构建活动，Gruber 在 1995 年提出的 5 条设计原则（Gruber, 1995）具有一定的影响力。

（1）清晰性：本体应该有效地表达所指定术语的含义，使之尽可能客观、完整，并以自然语言的方式将所给定义文档化。

（2）一致性：由本体得出的推论与本身的含义是相容的，不会产生矛盾。本体中定义的公理应该是逻辑一致的，如果从公理中推导出的句子同形式化定义相矛盾，证明该本体是不一致的。

（3）可扩展性：本体在设计中要考虑到这些共享词汇将来的应用范围，为一系列可能的任务提供概念基础，使本体的表达能被单调地扩展，即向本体中添加专用的术语时，不需要修改原有的定义。

（4）编码偏好程度最小：概念是基于知识层面而言的，而非依赖于符号编码系统，对不同的描述系统和表示方法，编码偏好最小化，要求采用便利的符号系统表示并尽可能最小化。

（5）最小本体承诺：对待建模对象给出尽可能少的约束，通过定义约束最弱的公理及提供交流所需的基本词汇，来保证本体的自由，能够按照需要进行本体的具体化和实例化，支持未来的知识共享活动。

2. 本体构建方法

应用目的不同，本体构建方法也不同，Mike 和 Michael 提出的骨架法（Mike and Michael, 1995）、Alexander 和 Steffen 的环形结构法（Alexander and Steffen, 2001）及 Noy 和 McGuinness 提出的七步构建方法（Noy and McGuinness, 2001）为很多研究者所采用。综合以上方法，我们认为构建具体步骤如下。

（1）识别目的和范围阶段，这个阶段需要弄清楚建立本体的原因及用途，以确定本体的使用范围及用户。

（2）建设本体阶段，包括本体的重用、捕获、编码化。重用现有本体，收集其他有用本体，利用已有的本体成果来扩充、修改现有本体，以减少本体构建的工作量。本体捕获主要识别相关领域中的关键概念和关系，进行精确无二义性的文本化定义，识别表达这些概念和关系的术语，并进行一致性检查。本体编码，对概念及概念关系进行形式化描述，确定类、属性、实例，选择描述语言进行有效性编码。

（3）评价阶段，采用一定的评估方法对构造好的本体进行质量评估。既包含