

# 论知识组织系统中的语义关系(上)

王知津 赵梦菊

**摘要** 本文从各类型语义工具作为一个系统的视角出发,分析了知识组织系统的概念、分类、功能与作用,阐述了语义关系的含义和类型,以及在知识组织系统中的意义,最后重点讨论了分类法、叙词表、本体、语义网络这四种语义工具中包含的语义关系。

**关键词** 知识组织系统(KOS);语义关系;分类法;叙词表;本体;语义网络

**分类号** G254

**Abstract** This paper analyzes the concepts, types and functions of knowledge organization systems(KOS), with a view of looking at various semantic tools as a whole system. In addition, the authors set forth the meaning and types of semantic relations and its importance in KOS. Furthermore, semantic relations contained in some semantic tools such as classifications, thesaurus, ontology and semantic network are emphatically discussed.

**Keywords** Knowledge organization systems(KOS); Semantic relations; Classifications; Thesaurus; Ontology; Semantic network

**Class Number** G254

## 1 引言

知识因子(节点)和知识关联(节点间的语义关系)是知识的两个要素,也是知识组织实践历来重视的内容。随着数字资源、网络资源的大量出现,以及人们对知识的需求越来越强烈,知识组织工作面临着更多的挑战和更高的要求,需要朝着自动化、智能化、集成化的方向发展,语义关系的深层次揭示和组织则成为关键。

语义关系是概念之间关联性的体现,它比语法关系更能体现出概念间的关系,在语义推理、知识组织系统互操作、信息检索等领域发挥着越来越重要的作用。比如,在信息检索领域,有的学者指出:“自然语言处理和语义关系尤其为 21 世纪的信息检索指明了前进的道路。”<sup>[1]</sup>能够表示知识结构体系及知识因子间的语义关联是知识组织系统需要具备的要素之一,知识组织实践的进步有赖于对语义关系的研究和应用,研究知识组织系统中的语义关系可以促进知识组织系统的进一步完善。

## 2 知识组织系统

网络知识组织系统/服务(Networked Knowledge Organization Systems/Services, NKOS)研究小组在美国计算机学会 1998 年的数字图书馆会议上最早使用了“知识组织系统”这个术语。随后提出

了它的定义:“试图涵盖用于组织信息和促进知识管理的各种类型的方案和体系。”<sup>[2]</sup>在 2008 年的 NKOS 会议上, Wright 将知识组织系统界定为:“一组概念的集合,有时也包含概念之间语义关系的描述。”<sup>[3]</sup>

国内比较广泛认同的定义是:“知识组织系统”(Knowledge Organization Systems, 或简称 KOS)是对人类知识结构进行表达和有组织地阐述的各种语义工具(Semantic Tools)的统称,包括分类法、叙词表、语义网络、概念本体以及其他情报检索与标引语言。<sup>[4]</sup>

实际上,“知识组织系统”由来已久,如传统的分类法、叙词表等,只是未使用“知识组织系统”这个术语来统称这些知识组织工具。随着越来越多的知识组织工具不断涌现,如本体、语义网络等,各种工具之间的界限越来越模糊,集成和互操作也成为发展的趋势,使用“知识组织系统”最能囊括所有这些促进知识组织的语义工具。本文所使用的“知识组织系统”是指各种能够组织知识因子、揭示知识关联(语义关系)、阐述知识结构的语义工具。

### 2.1 知识组织系统分类

关于知识组织系统的分类,先后有过多种不同的看法。

NKOS 网站将知识组织系统划分为三种主要类型:①词汇列表(规范文档、术语表、地名辞典、字典);②分类与归类(主题词表、体系分类表、

归类表和知识分类表);③关系模式(叙词表、语义网、知识本体)等。<sup>[5]</sup> Hodge 依据知识组织系统的复杂度、结构、用语关联性以及功能等特点,把知识组织系统分为 3 个层次、11 种类型:词汇列表(规范文档、术语表、字典、地名辞典)、分类表(标题表、分类表)以及概念一览表(叙词表、语义网络、本体)。<sup>[2]</sup> Stock 等人把知识表示中的知识组织系统分为五种:术语表、分类体系、叙词表、本体,以及作为知识组织系统边缘的大众分类法。<sup>[6]</sup>

Hill 等根据概念关系结构的强弱和对自然语言控制的程度,将知识组织系统分为词单类(可选词单、词汇/字典、同义词环)、分类与归类类(图书分类法、知识分类表)和关联组织类(知识本体/实用分类法、叙词表等)三种类型。<sup>[7]</sup> 马文峰、杜小勇根据概念和关系的揭示程度,将知识组织系统划分为概念类聚系统和概念关联系统两个层次。<sup>[8]</sup>

大多数学者都依据各种知识组织系统在语义关联上的功能和特点对其进行划分,从它们产生的环境和发展历程上看,知识组织系统中又存在着“传统型——过渡型——新兴型”这三个特点的检索语言和语义工具。所以我们可以将其划分成三种广义的知识组织系统:第一种是来自图书情报等领域在知识组织实践中产生和发展的传统知识组织系统,如分类法、叙词表、主题标题表、规范文档等;第二种是传统知识组织系统在网络环境中的延伸和发展,如分类法、叙词表、主题标题表、规范文档等的网络应用;第三种则是在网络环境中产生和发展起来的语义工具,如本体和语义网络等。在具体应用中,我们可把知识组织系统泛指为各种传统的或新型的语义工具,如各种分类法和叙词表、军事领域本体、UMLS 语义网络等。

## 2.2 知识组织系统的功能与作用

出于不同的知识组织目的可选择构建不同类型的知识组织系统,它们在复杂性、结构、描述语言、规则上存在很大差异,但都在知识组织实践中发挥着重要的功能和作用。

Hill 等认为,知识组织系统所支持的功能包括:①描述:受控的标识集,用以描述事物;②定义:标识的意义;③翻译:等价表达形式之间的匹配;④导航:在一个有组织的表达结构中的联接。<sup>[7]</sup> Shiri 认为知识组织的主要作用是:①使用叙词表与分类法用于各种数字资源的跨库浏览与检索;②利用现有的叙词表创建数字图书馆的本体模型;③分类法与受控词汇为数字资源提供了通用的知识表达工

具;④使用知识分类表描述数字资源的不同层面,为不同数字仓库提供统一的、有序的检索。<sup>[9]</sup>

王军认为知识组织系统具备五大功能:①分类法和主题词表为数字图书馆资源的组织和管理提供了一个知识框架;②为数字图书馆业已积累的元数据资源提供有效的利用手段;③为用户提供了一个检索、服务、教育一体化的知识空间;④一个自丰富、自增强、自适应的知识系统;⑤实现数字图书馆的知识管理的技术基础。<sup>[10]</sup>

以上学者对知识组织系统的分类均偏重传统知识组织系统在数字图书馆环境中的功能和作用。根据本文对知识组织系统的定义和类型划分,知识组织系统的功能和作用早在传统的分类法、叙词表等服务于知识组织实践时就开始了。广义上的知识组织系统服务于人类知识组织实践的全过程,它们通常都具备规范、描述、定义、模型化知识结构、提供导航等基本功能和作用,其实现依赖于具体知识组织系统在信息组织实践中的应用。

## 2.3 几种典型的知识组织系统

文献分类法和叙词表是传统知识组织实践中最主要的两种语义工具,随着知识组织实践的发展,对知识因子和知识关联的揭示更加深入,本体、语义网络等的应用日渐广泛,成为研究的热点。各种知识组织系统各有用武之地,都是不可或缺的。本文选取其中的几种进行简单讨论。

### 2.3.1 分类法

分类法是一种重要的传统的知识组织系统。它以资源内容属性为基础,主要用分类号、学科或专业类目来完整系统地表达知识因子,通过隶属、并列和设置交替类目、多重列类、类目参照等方式揭示知识关联。分类法的应用十分广泛,在文献排架、网站架构、数据库导航等方面,对概念和主题进行层层划分,展现其间的等级关系。国内外比较著名的分类法有《杜威十进制分类法》(DDC)、《国际十进制分类法》(UDC)、《国会图书馆分类法》(LCC)、《中国图书馆图书分类法》等。

传统分类法是一种比较成熟的知识组织方法,其良好的等级结构、知识体系以及类目关系使得它们在数字资源、网络资源中得到了广泛应用。目前,网络上的信息组织分类主要采用两种方式:一是沿用传统的分类方法,主要以学术资源为主,比如,以图书馆学情报学为特色的综合学科信息网站 BUBL(<http://bubl.ac.uk/>)使用杜威十进制分类法(DDC)提供学科分类导航。二是采用自

编的分类系统,这在门户网站上较为常用。<sup>[11]</sup>比如,门户网站雅虎(<https://www.yahoo.com/>)采用自编分类法,使用16个大类分类组织网络信息资源。

### 2.3.2 叙词表

叙词表也是一种重要的传统的知识组织系统。叙词表又称主题词表,它吸收了元词法的后组配原理、标题法的参照系统、分面组配分类法的分类规则等优点,在多种知识组织系统的基础上发展而成。叙词表从资源主题出发表示知识因子,并使用受控(或优选)词来规范知识因子的表达形式,使用参照系统和索引系统揭示知识关联。中国国家标准《汉语叙词表编制规则》(GB13190—915)将叙词表定义为:“叙词表是将文献、标引人员或用户的自然语言转换成规范化语言的一种术语控制工具;它是概括各门或某一学科领域并由语义相关、族性相关的术语组成的可以不断补充的规范化的词表。”<sup>[12]</sup>

叙词表的范围很广,既有助于通用广泛领域的综合性叙词表,也有用于专门狭窄领域的专业性叙词表。国内外比较著名的主要叙词表有《美国国会图书馆标题表》(LCSH)、《医学主题词表》(MeSH)、《工程与科学主题词表》(TEST)、《汉语主题词表》等。叙词表的词汇控制方法和参照系统较为完善,词间关系指示明确,在网络知识组织中也较为常用,各大主流数据库都在不断改进叙词表在其知识组织中的作用。网络环境下应用叙词表的主要形式是使用已有印刷型叙词表或重新编制数字化叙词表。比如,Intute 学科信息门户(<http://www.intute.ac.uk/medicine/>)使用《医学主题词表》(MeSH)组织医学资源;国际农业数据库 AGRIS (<http://aims.fao.org/standards/agrovoc>)使用多语种叙词表 AGROVOC。

### 2.3.3 本体

本体(ontology)最初是哲学中的一个概念,用来解释存在的本质和规律。知识组织系统中的本体则是语义网中信息组织的核心概念,用来描述一群人所共有的针对一个特定领域的概念框架或知识,或者说是这种概念框架的形式化表达。Gruber 于1993年给出的定义最为流行:“本体是概念化的明确的规范说明。”<sup>[13]</sup> Studer 于1998年进一步将其完善为:“本体是共享概念模型的明确的形式化规范说明。”这揭示了本体的四层含义,即:(1)概念模型(conceptualization),指通过抽象

出客观世界中一些现象的相关概念而得到的模型,概念模型所表现的含义独立于具体的环境状态;(2)明确(explicit),指所使用的概念及使用这些概念的约束都有明确的定义;(3)形式化(formal),指本体是计算机可读可处理的;(4)共享(share),指本体中体现的是共同认可的知识,反映的是相关领域中公认的概念集。<sup>[14]</sup>

本体实际上是一种用 is a、部分—整体、属性—值等关系及公理等规则联系起来的观念层次体系。具体地说,本体就是一整套对某一领域的知识进行表述的词和术语,编制者根据该知识领域的结构将这些词和术语组成等级类目,同时规定类目的特性及其之间的关系。<sup>[13]</sup>

本体有多种类型,它们的差异取决于所在的知识领域、具体应用、描述框架和采用的构建方法与工具等。例如,由普林斯顿大学开发的面向词汇处理的 Wordnet,它是基于心理语言规则的英文词典,以 synsets(特定的上下文环境中可互换的同义词的集合)为单位组织信息,通常被认为是词汇或术语本体。又如,由爱丁堡大学人工智能应用机构(AIAI)开发的服务于企业决策的企业本体(enterprise),它结构化了关于企业的知识条目及定义,帮助企业进行知识的获取、表示和操作,是一种应用型本体。

### 2.3.4 语义网络

语义网络的概念最早由 Collins 和 Guillian 提出<sup>[15]</sup>,它是一种用来表达概念间相互关系,揭示深层语义信息的知识组织系统。一个语义网络是某个领域知识或概念之间关系的网式图,一般由节点和有向链或有向弧构成。其中,节点一般划分为实例节点和类节点两种类型,表示问题领域中的物体、概念、时间、动作或者态势等,节点之间带有标识的有向链或有向弧表示节点之间的语义关系。图1是“手机”语义网络的一个片段。

语义网络不但能同时生成和表示知识因子和知识关联,而且其描述的对象或范围也比本体广泛。由于语义网络表示知识具有简洁直观的特点,并且与人类联想记忆的思维模式相符合,因此在自然语言处理和专家系统建设等领域得到了广泛的应用。

一般来说,传统知识组织系统的处理对象主要是文献资源,面向的对象主要是人;网络环境下知识组织系统的处理对象主要是电子资源,面向的对象主要是机器。为了使机器理解人所理解的

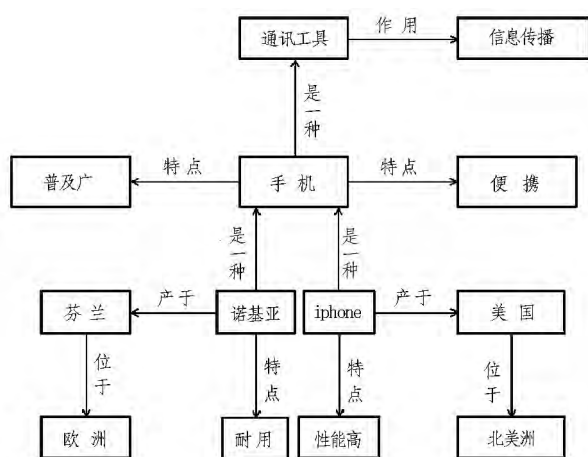


图1 “手机”语义网络片段

知识关联,实现自动推理等功能,需要在这些新型语义工具中揭示知识因子间更为深入细致的语义关系。

### 3 语义关系

语义关系对于展现知识因子之间的相互联系具有重要作用,最早定义和研究语义关系的是语言学、逻辑学、心理学等领域。信息组织中有关语义关系的研究集中在概念和术语方面,随着研究的深入,目前正在向识别、处理和关系管理转变,在信息提取、信息检索和自动摘录中有所应用。<sup>[1]</sup>

#### 3.1 语义关系的含义

知识组织中的语义关系是指知识因子中的知识关联。知识因子在不同语义工具有不同的表现形式。在分类法中,知识因子是学科/专业类目词;在叙词表中,知识因子是概念主题词;在大众分类法(folksonomy)中,知识因子是用户标签(tags);在语义网络中,知识因子是概念、事件、过程、状态等。对知识因子中语义关系的揭示是知识组织的前提之一。

从语义关系的一些公式表达来看,简单的语义关系可由[概念1]→[关系R]→[概念2]这种三元组的形式表现,比如,[衣服]→[来源于]→[制衣厂]。也可表示为二元关系R(概念1,概念2)形式表示,比如,来源(衣服,制衣厂)。其中“[关系R]”和“R”表达了概念1与概念2间存在的语义关系。

从语义关系表现的对象来看,它可以指思维中概念间的关系(conceptual relations)、词汇之间的关

系(lexical relations)或者文本段之间的关系。<sup>[16]</sup>

不同的学科对于语义关系持有不同的观点,可以把这些观点的不同看成是学科间相互区别的原因之一。语言学将语言看成一个系统,系统的一个重要特点是它内部成分之间的相互联系性,其中,词汇由于语义上的相似或相同关系而构成词汇—语义系统,研究词间的这些关系并揭示其特征和规律是词汇—语义系统研究的一个重要内容。逻辑学在逻辑层面对语义关系进行分析,认为语义关系在人类的理解和推理中具有重要作用。Chaffin和Herrmann等心理学家把对语义关系的理解看成是一种心理现实,通过分类实验证明人类可以判断相似关系、区分不同关系、识别一般关系的实例。<sup>[17]</sup>在信息科学领域,语义关系常用来指依据语义信息而建立的关系,它揭示、描述信息之间的内在联系,是客观知识体系框架的重要构成因素。<sup>[18]</sup>

#### 3.2 语义关系的类型

对语义关系进行分析和分类有利于理解语义关系的特性,帮助我们识别知识组织系统中各种语义工具所包含的语义关系类别。许多学者根据语义关系的含义和性质分别从语言学、心理学、逻辑学和计算机学角度来认识和研究语义关系,并尝试以某种特定的方式来划分语义关系的类型。

Landis等识别出了反义、同义、类包含、部分整体和事件5个语义关系。<sup>[19]</sup>其中,类包含(class inclusion)是指一个类中的所有类都包括于另一个类,比如“所有植物都是生物”,“植物”和“生物”就是类包含关系。随后,Chaffin和Herrmann提倡语义关系的精编和分裂,并把31种语义关系划分为5组:对比、相似、类包含、事件关系、部分整体关系,为区分多种多样的语义关系提供了的标准。<sup>[17]</sup>大多数学者如Peter和Weller等都把语义关系分为纵聚合关系和横组合关系。<sup>[20]</sup>纵聚合关系是出现在相同位置的词或短语之间的垂直关系,通常属于同一个语义场,它独立于文本,是在KOS中已经建立或规定的关系,具体包括等同、等级和相关关系。横组合关系主要指同一语句或文本中词的搭配同现关系,通常不固定。

除对术语或概念进行定义和组织外,表达和显示术语/概念间的语义关系是各类知识组织系统的另一重要内容。通过知识因子间等级、等同语义关系的揭示,知识组织系统才得以在知识的有序存储、有效检索等知识组织实践中发挥作用。

不同类型的知识组织系统在语义关系的数量和类型、强弱程度、复杂程度以及对其界定和表达方式上存在着或多或少的差异,有着自身独特的语义关系体系,故不便于使用同一个语义关系类型的划分标准对其割裂。本文拟保留各知识组织系统中原有的语义关系体系,仅对上文提出的几种语义工具中的语义关系进行分析。

#### 4 分类法中的语义关系

分类法中的语义关系显示得比较直接和明确,本文以《中图法》为例,分析其中的等级关系、等同关系、相关关系、并列关系等语义关系。

##### 4.1 等级关系

作为一种等级列举式分类法,等级关系是分类法所有的类目关系中最清晰和明确的关系,等级关系在分类法中又称作上下位类关系。《中国图书馆图书分类法》第四版规定类目体系中的上下位类之间的关系包括属种关系、整体与部分关系、全面与某一方面的关系。分类法利用上下位类的设置实现等级上的聚类。

在具有等级关系的类目中,被划分的类称为上位类,即母类;区分出来的类被称为下位类,即子类。上位类必定能包含下位类的外延,下位类必定有上位类的属性。上位类和下位类是相对的,除了一级类目和最低级类目外,其他级别的类目既可以是上位类又可以是下位类,例如,“初等数学”是“初等代数”的上位类,同时又是“数学”的下位类。以基本类目为起点,各级类目以等级的形式层层展开为一级类目、二级类目、三级类目等,实现知识结构的等级式组织。《中图法》以字母、数字、符号等表示类目间的等级关系。例如,

G 文化、科学、教育、体育

G2 信息与知识传播

G25 图书馆事业、信息事业

G250 图书馆学、情报学

值得注意的是,上下位关系并不等于严格意义上的概念等级关系,在“全面与某一方面的关系”中,方面关系指学科及其问题、事物及其属性等之间的关系。虽然在应用时能够便于检索者扩检和缩检,但实际上只能归属为“准等级关系”。

##### 4.2 等同关系

分类法中的等同关系表示两个概念具有相同的内涵和外延,即同一词汇的不同表达形式,又称

“交替关系”。交替关系指交替类目与其使用类目之间的关系,通常在一些知识门类具有多重从属关系,同时隶属于两个或两个以上的学科或部门时使用。<sup>[21]</sup> 交替类目的设置在很大程度上揭示了概念主题间的关系,实现两个等同概念的聚类,保证了同一内容文献的集中以及知识体系的完整性。

交替类目是一种供使用者选择的类目,与使用类目之间是等同关系,用[ ]表示使用时不使用该类目而使用其正式的类目。在手工检索中,交替类目一般以参照的方式指向使用类目;在机检中,通常以链接的方式指向交替类目。例如,

[S17]农业地理学 宜入 F319.9 农业经济地理

##### 4.3 并列关系

并列关系表示同一个上位类划分出来的、在外延上不相交、处于同等地位的一组类目之间的关系,这些类目之间互称为同位类。同位类间具有相互联系的特点,因为它们由同一个上位划分出来,带有上位的属性或被上位限定;另一方面,他们之间在外延上互相不相交,因此又是互斥的。例如,

G7 各类教育

G71 职业技术教育

G72 成人教育、业余教育

G74 华侨教育、侨民教育

G75 少数民族教育

##### 4.4 相关关系

相关关系指类目之间除等级、等同、并列等关系外的其他联系,表示类目的含义相关,可以认为是一种交叉关系,例如“图书馆学和情报学”。相关关系通常用“参见”的方式来表示,参照类目的设置反映了类目体系中被分散了的联系,与交替类目一起,在很大程度上保证概念的集中和知识体系的完整性。

类目参照一般用于内容联系具有揭示价值,但在分类体系中被分散了的类目之间,通常采用互逆的方式在相关门类下注明。<sup>[21]</sup> 例如,

TG39 高能成型 参见 TF124.36 高速成型

TF124.36 高速成型 参见 TG39 高能成型

在上述四种语义关系类型中,等级关系、并列关系是揭示知识因子间关系的主要线索,反映了知识因子的纵向关联;等同关系、相关关系则是对知识因子间主要关系的补充,揭示被类目体系分散了的横向关联。通过两者的结合,分类法在知识因子间建立了纵横交叉严密实用的知识关联。

(未完待续)