

计算思维视角下信息组织核心概念 与描述模型的再认识*

陈 翀 张婧莹 李春秋

(北京师范大学政府管理学院, 北京 100875)

摘 要: [目的/意义] 大数据时代, 信息的规模和复杂性成为信息组织面临的必然挑战。为了适应这样的挑战, 并顺畅地将人类组织信息知识转换为计算机可理解的形式, 本文拟探讨信息组织理论中蕴含的数学和工程思想, 对核心概念进行形式化定义, 帮助提升信息组织知识表述的科学性。[方法/过程] 选取信息组织的核心概念, 从离散数学和面向对象建模中寻找具有共性的内容。基于计算思维, 比较了面向对象中的核心概念和信息组织中类目构建、元数据定义等内容的相通之处, 从集合划分的角度来解释概念划分与分类体系构建, 用二元关系来表达叙词表的词间关系, 并用图论来概括信息组织的逻辑结构。[结果/结论] 本文建立了面向对象的信息组织描述模型, 揭示了概念划分、分类体系构建、词间关系的数学内涵, 分析了三种逻辑组织结构的特点。本研究的创新之处在于用计算思维对信息组织中的定性知识加以抽象, 推动核心概念的精确化表示, 不但为利用相关数学概念和性质来完善信息组织知识体系提供基础, 还探索传统的定性理论与现代技术之间的更好融合, 以巩固信息组织理论体系的一致性。

关键词: 信息组织 计算思维 面向对象 描述模型 词间关系 分类体系构建 信息组织逻辑结构

分类号: G350

DOI: 10.31193/SSAP.J.ISSN.2096-6695.2021.04.02

0 引言

信息技术的进步和大数据时代的到来, 推动信息组织在对象、方法及应用环境方面逐渐步入多元化。在资源数量大、形式多、用户需求灵活变化的情况下, 需要传统组织理论与计算机技术相互协同才能提高信息组织的效果和效率。图书情报领域的学者们已经意识到, 为了适应新的挑

* 本文系国家自然科学基金项目“面向科研人员定量评价的多维学术专长识别及属性度量研究”(项目编号: 21BTQ065)的研究成果之一。

[作者简介] 陈翀 (ORCID: 0000-0002-9704-1575), 女, 教授, 博士, 研究方向为知识组织、文本挖掘, E-mail: chenrong@bnu.edu.cn; 张婧莹 (ORCID: 0000-0001-8231-2064), 女, 硕士, 研究方向为知识组织; 李春秋 (ORCID: 0000-0003-1520-1297), 女, 讲师, 博士, 研究方向为政府数据管理、元数据, E-mail: lichunqiu@bnu.edu.cn。



陈翀, 张婧莹, 李春秋. 计算思维视角下信息组织核心概念与描述模型的再认识 [J]. 文献与数据学报, 2021, 3 (4): 012-025.

战, 信息组织需要建立新型的数据表示方式^[1], 更新知识体系^[2], 充分揭示信息特征与信息之间的内在逻辑联系^[3], 以满足用户对信息组织的细粒度、关联性要求。可以说, 信息组织在以文献为主要对象的 1.0 时代发展到以知识组织为特征的 4.0 时代^[4]的过程中, 吸纳了多个学科分支的理论方法。这让我们思考, 为了使信息组织知识体系更具系统性、一致性, 来自不同学科思维模式的理论方法应从哪里融合? 这样的思考也是在呼应很多初学者的疑问。不少信息组织课程的学习者反馈课程内容庞杂^[5], 理论之间的关联性不强。

本文的思路是: 从概念定义的层面把经典的组织理论与来自计算机领域的技术方法建立密切联系。对以往基于文献组织的定性描述进行抽象提炼, 找出其数学上或工程上的对应知识, 建立信息组织核心概念的形式化定义。这不但能使信息组织知识脉络更加一致, 还有助于将形式化表述的信息组织方法更容易地转换为系统实现, 扩展传统理论对现实应用的指导能力。

由于信息组织涉及到的理论、方法和技术覆盖面很广, 因此, 为了甄别核心知识点, 本文用经典信息组织教材内容进行对比分析, 在概念逻辑原理、分类法、主题法、信息的描述标引等主干模块中各选择一个基础单元, 运用计算思维进行阐释。具体如下: (1) 在概念逻辑原理中重点表述概念的划分, 将概念划分与集合论中集合划分对应起来, 使构建分类体系有了数学基础; (2) 对于主题法, 重点探讨叙词表的词间关系, 从二元关系的角度进行对应; (3) 在描述标引中, 立足于“资源是对象”这一要点, 用面向对象的通用思想建立信息组织的描述框架; (4) 对于资源组织的逻辑结构分析, 本文借鉴图论和数据结构归纳了三种组织结构的特点。上述探索的意义在于, 为利用相关数学性质对定性知识进一步引申扩充提供了条件, 有助于完善信息组织知识体系。例如, 将词间关系定义为二元关系后, 在关系运算支撑下可以丰富词间关系类型等。

本文第一节为相关研究; 第二节是基于面向对象的信息组织描述模型; 第三节是用集合和二元关系对概念划分、分类体系的类目划分和词间关系的解释; 第四节用图论知识探讨信息组织的逻辑结构, 并阐述各自的优缺点; 第五节分析本文对课程教学概括信息组织知识体系的价值; 第六节是总结。

1 相关研究

1.1 计算思维视角

按照中国大百科全书第三版的词条定义, 信息组织是揭示和描述信息的内容与外部特征, 按特定顺序排列并依据一定结构展示, 从而方便用户对信息的获取和利用^[4]。信息组织包括对信息的获取、存储、筛选、描述、分类与分析利用全链条上的任务。贾君枝教授根据数据网络的需要, 将信息组织的处理功能分三层, 即资源层、元数据记录层、词表层^[6]。它们分别对应着对资源的采集整理、描述标引和基于资源集合的概念与分类体系构建。按功能分层的一个好处是容易分清不同层级涉及哪些任务。总体来说, 由于全链条长、资源类型不一、用户需求灵活变化等原因, 信息组织可以视为一系列复杂任务。知名华裔学者周以真教授于 2006 年提出“计算思维”^[7], 认为这是人类求解问题的一种有效途径, 它采用抽象和分解来应对庞杂的任务或设计巨



大而复杂的系统。

计算思维是指在问题求解、系统设计以及人类行为理解等思维活动中运用计算机科学的基础概念。它体现了数学思维和工程思维的互补与融合,并且不仅适于计算机领域,还适于现实生活中的大量问题。计算思维的一个特色是利用不变的量简明扼要地刻画系统的行为,进行多个层次的抽象,这样就无需一次解决所有细节^[7]。举例来说,语义网堆栈(semantic web stack)^①的七层模型就体现了用这种思维方式来描述语义网涉及到的多种处理任务。对信息组织的描述也可以借鉴这种方式。

形式化描述是计算思维的一种表现。在软件工程中,是指用精确的数学逻辑描述方法指导软件的开发;在更宽泛的领域,是指用相对完备、无歧义的方式描述事物的行为特征或属性。表达某个系统的形式化方法有多种,例如通过明确定义状态和操作来建立该系统的模型;又如通过代数方式、逻辑表达式或图的方式来定义这个系统必须要满足的一些性质等。对信息组织中明确而稳定的内容就可以借鉴这样的方式加以描述。一个典型的例子是定义本体,一般是遵循形式化规范来表达领域中各种知识,包括对实体、类、关系和限制的定義等。例如 owl 语句“EquivalentClasses (:Parent ObjectUnionOf (:Mother :Father))”就是用类似数学表达式的方式定义了类别 Parent 与 Mother 和 Father 并集的等价性。

由于信息组织的任务是直接或间接围绕资源展开的,因此,如果把现实世界看成问题域,把问题域中的客观事物抽象表示为对象,将其静态特征定义为属性,将其行为定义为操作,就可以用面向对象的方式阐述并解决问题。面向对象也符合计算思维的特点,是以认识论为基础的一种对现实世界的理解和抽象方法。不仅限于程序设计,在数据管理、人工智能等领域也有广泛应用。相比面向过程的方法,它以对象为各类系统的核心,注重保持事物及其相互关系的本来面貌,更适合用于复杂事物的管理。事实上,它在信息资源组织中早有应用先例。1992年,美国国防部高级研究项目局(ARPA)支持计算机科学技术报告项目(The Computer Science Technical Report, CS-TR)设计一种适于数字图书馆的基本架构。1995年KAHN R和WILENSKY R的报告^②中将数字对象模型描述为,数字对象中包括标识符、属性、内容、签名和日志^[8]。也就是说,该模型将数字资源的内容和其外部特征定义在一起构成一个对象,这样的统一描述屏蔽了不同数字资源的差异,并将对象的属性和操作记录与其内容集成以便于管理,只是这一模型没有涉及组织资源要进行的操作。

1.2 核心概念的选取

为了选择信息组织核心概念,本文对国内外主要教材^[9-13]进行了比较,发现虽然有的以传统的、面向文献的组织理论及方法为主,有的扩充了网络信息组织、web 2.0 信息组织等方面的内容,但这些原理在信息组织的基本原理上有较普遍的共识。具体包括语言学原理、知识分类原理、概念逻辑原理和系统论原理,它们构成了分类法、词表及检索系统的基础,如表1所示。表1第一列是依照教材总结的经典内容模块,第二列是各模块涉及的理论性知识,即不含模块中的操作方法、构成情况等介绍性内容或易发生改变的内容,第三列是这些理论性知识所依据的基本原理,第四列是本文对基本原理进行解释和扩充所选取的角度。需要说明的是,这些经典信息组织的基本原理取自语言学、哲学、逻辑学、系统论等不同学科领域,相互之间并不密切承接,这也是整个知识体系内容显得散乱的一个原因。



陈翀, 张婧莹, 李春秋. 计算思维视角下信息组织核心概念与描述模型的再认识 [J]. 文献与数据学报, 2021, 3 (4): 012-025.

表 1 信息组织经典内容的基本原理

内容模块	主要理论性知识	基本原理	本文探讨的角度
信息组织基本原理	信息组织的任务界定、原理介绍	语言学原理、知识分类原理、概念逻辑原理、系统论原理	集合论、图论、面向对象方法
分类法	分类的概念、类目体系构建原则、类目关系类型	知识分类原理、概念逻辑原理	集合划分、二元关系、图论
主题法	主题的概念、叙词词间关系、概念组配原则	知识分类原理、概念逻辑原理、语言学原理	等价关系、偏序关系、其他二元关系、集合运算
信息标引与描述	主题因素与关系、主题概念分解与概括、元数据	知识分类原理、概念逻辑原理、语言学原理、属性定义	二元关系、面向对象描述

2 信息组织的描述模型

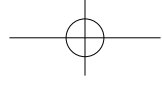
信息组织可视为围绕资源对象的一系列操作，有些是在对象内部、对属性的直接操作，有些是在对象外部、利用多源信息对单个对象或对象集合进行的操作。本节用面向对象的思想建立信息组织的描述模型，阐述信息组织中蕴含的继承、多态、封装、复用等概念。通过对比相通之处，来探讨该模型对信息组织抽象的合理性，从而试图用统一的框架将相对稳定的组织思想从变化的实现方式中分离出来，用易于扩展的方式描述信息组织的主要任务。

2.1 基本概念的对对应关系

面向对象中，属性和操作以对象为中心，操作依附于对象。对象是稳定的，而对象的行为是可变的。在信息组织的问题域中，资源对象的属性对应了其静态特征，操作是为了适应筛选、描述、排序、利用等组织活动而需要进行的属性设置、访问及处理。信息组织中，资源对象表现出稳定性，因为其本质特征并不随应用场景改变而发生变化，但资源关联的操作需要适应不同场景的组织活动的变化。表 2 梳理了信息组织与面向对象中的基本概念之间的对应关系。

表 2 信息组织与面向对象基本概念的对对应关系

面向对象	信息组织
对象	资源
属性	静态特征，对应基本元数据项。
操作	对资源特征的赋值、访问、比较、计算等，用于支撑对资源的筛选、描述和排序等组织活动。
对象是唯一的	资源个体具有唯一性。
类	具有共性的资源可被抽象成类、具有共性的子类也对应上层类。
继承	对有共性的类别进一步抽象概括为上层类，上、下层的类通过共享特征形成分类体系。下层类既有上层类的特征，也有自身特征。
多态	同种操作为适应不同应用或资源类型而进行改写。



2.2 面向对象的信息组织描述模型

设定资源集合 R 。归纳该类资源的特征，记为属性 $\langle a_1, \dots, a_i, \dots, a_n \rangle$ ，对任意实例 r ($r \in R$)，这些属性的值不同。与组织活动有关的内部、外部操作，分别记为 op_j 和 op_g 。内部操作包含在对象 r 中，可以进行属性获取、赋值等处理，配合外部操作完成组织活动，如分配资源 ID、建立资源间的关联等。外部操作可以对多个对象的属性进行处理，如排序、比较、计算等，还可以删除对象，达到资源筛选的目的。实现中，这些操作就是函数定义的基础。面向对象的信息组织描述模型如图 1 所示。

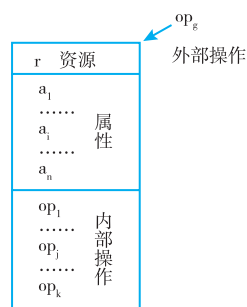


图 1 面向对象的信息组织描述模型

除了资源，信息组织中其他独立事物也可以抽象为对象，如资源的载体、标准规范等，本节仅以资源为例。

2.2.1 类和实例

自底向上地看，类是对象集合的抽象概括，对象的共性特征构成了类的特征，每个对象是该类的一个实例。信息组织中的分类实际上是依照资源某项特征的取值而对资源集合 R 的划分，得到的资源子集用类别标签概括。例如，文献集 R 中的对象有属性“作者”，记为 a_i ，取 $a_i = \text{“鲁迅”}$ 会得到一个对象子集 R' ，其类别标签可记为“鲁迅文集”。分类的前提是具有预先定义的类别体系，层级体系是最常见的，这种层级就是对资源集合或类别集合中的共性不断抽象概括的结果。类别之间的共性特征可以用更宽泛的类别来涵盖，如“水墨画”“油画”可以纳入“绘画作品”的类别。这与面向对象中，父子类之间通过继承形成层级关系是一致的。

2.2.2 继承和多态

面向对象中，类的一个重要特点是具有继承性，即允许子类继承父类的特征，子类只需要定义自己特有的属性及操作。在信息组织的分类中，经常遇到一些类别兼具多个类的特点，例如，类别“古籍”继承自“书”和“需要特殊保护的物品”两个类。类的继承性使分类体系减少了知识描述的冗余，保障了类别在细化过程中子类与父类本质特征的一致，也使分类体系的上层具有通用性、下层具有特殊性。当多个类的属性和操作具有共性时，可以定义一个单独的类 A 表现这些共性特征，这样 A 的子类 B 就能继承这些特征。

多态是指子类除继承父类的属性和操作，还可以定义自己的操作来代替其任何超类的同名操作，这确保了对具体事物具体分析灵活性。信息组织中，由于资源的产生和使用都有各自的环



境, 所以进行同样组织活动的操作方式并不唯一。例如对纸质图书和网页这两种信息资源的分类操作就不相同。假设“信息资源”类定义如图 2, 有属性 a_1 和 a_2 分别记录标题、摘要信息, op 是某种分类操作, 它将提取标题和摘要, 参照某分类体系标引资源的类别。假设“信息资源”有“纸质图书”和“网页”这两个子类, 子类“纸质图书”的分类操作与父类中的定义相比, 细化了向分类体系映射的规则; 而子类“网页”的分类操作则是直接利用属性的内容与文本处理的相关算法进行自动分类, 因为网页资源的内容是数字化的, 且数量很大, 所以适合这样的处理。可以说, 继承和多态保证了组织不同类型资源时, 目标上的一致性和实现上的灵活性。

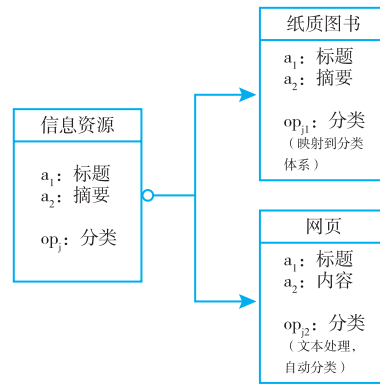


图 2 信息组织活动中体现出的多态性举例

2.2.3 复用和封装

信息组织中离不开标准和规范, 它们的作用是将属性定义或操作方法提炼并固定, 以便于用到其他资源的组织上, 这与面向对象中复用 (reusable) 的思想不谋而合。事物的属性和操作封装在一起作为一个整体使用是可复用的基础。复用有多种类型, 图 3 展示了信息组织中体现聚合复用的例子。聚合复用是将一个对象纳入到其他对象中作为成员使用, 该对象的内部实现细节对其他对象是不可见的。图 3 中, 图书的属性之一是载体形式, 载体有纸质和电子版之分, 不同载体形式有自己独特的属性, 如纸质载体的尺寸、电子载体的分辨率等。如果将载体形式单独定义并作为图书的成员属性, 这样对“载体形式”的定义所做的修改、完善也会体现至图书中, 并不必修改图书的其他属性和操作。

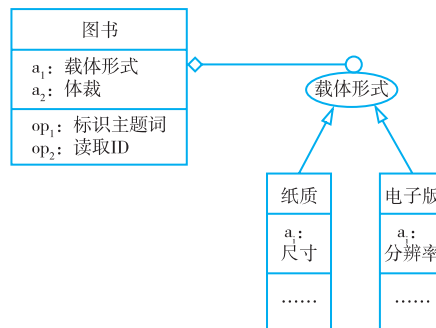


图 3 信息组织中的复用思想

封装是指将事物的属性和操作作为一个整体对待。其好处是隐蔽事物的内部细节，只是向外部提供访问接口，降低了对象间的耦合度，减少了修改引起的“波动效应”。从程序设计角度看，封装是外部不能随意存取对象的内部数据，对象内部的修改对外部的影响很小；从系统设计的角度看，封装能简化系统设计、使各部分结构清晰。语义网堆栈就借鉴了开放系统互连参考模型（open system interconnect, 简称 OSI）的层次设计思路，编码层和资源描述层、逻辑规则层等相互分开，体现了逐层封装。在信息组织中，很多活动并不在同一个层次上。封装有利于突出主要目标，保持实现的灵活性。除此以外，在不同的信息系统之间经常需要互操作，封装也是其实现基础。

2.3 面向对象方式描述的信息组织实例

图4是对信息组织从面向对象角度定义的一个例子。book是一个类，r、r'是两个具体的对象，继承了book类的共性特征。author类继承自父类“people”，有“id”“姓名”等属性，也定义了“研究方向”这一新的属性。在操作上包括对属性的赋值 op_1 ，按照给定属性的值进行排序 op_2 ，以及输出属性的值 op_3 ，还有一些外部操作，对资源集合进行处理。该例对资源的组织重点展现了描述和利用两种活动。“描述”是通过一个外部操作 op_g' 将特定的作者对象u赋予书的对象r。“利用”主要是反映了按关键词检索，根据给定的关键词t，通过外部操作 op_g 从资源集合{r, r'}中筛选主题为t的个体。为了简化说明，图中将样例操作作用函数调用的方式表示。

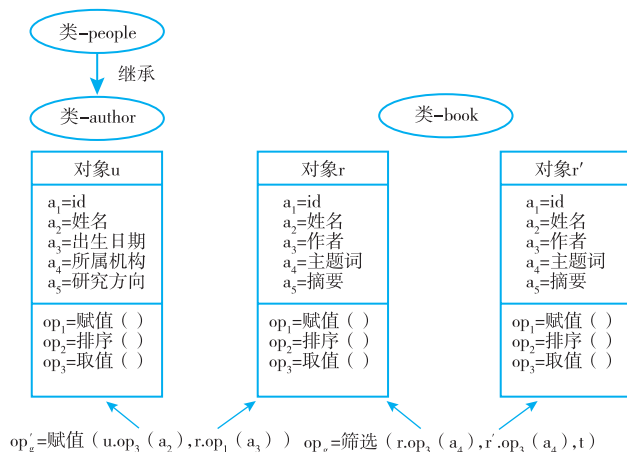


图4 用面向对象方式描述的信息组织实例

信息组织中的很多概念和经验做法与面向对象的思想吻合。而面向对象是解决问题的一种通用思路，在理论和应用上都比较成熟。用面向对象的思想完善信息组织的统一描述框架及其中涉及的信息组织活动，更容易被转换为计算机可理解的方式，能促进自动化组织系统的设计和实现。通过定义对象的属性及操作还可以将信息描述和信息组织方法联系起来。在此基础上，文献、互联网网页、数据集等各类型资源的具体组织方法可定义为一个统一框架的不同实例化形式。

用面向对象方法进行大型、复杂事物的描述与任务定义已在数据管理和程序设计应用中得



陈翀, 张婧莹, 李春秋. 计算思维视角下信息组织核心概念与描述模型的再认识 [J]. 文献与数据学报, 2021, 3 (4): 012-025.

到广泛的检验。目前图书情报领域对信息的描述、分类、检索等加工利用已越来越依靠计算机完成。建立统一的信息组织描述模型, 使组织策略容易转换为机器可理解的形式, 同时这样的方式也兼容知识组织、语义网等细粒度、深层次的信息组织发展要求。

由于信息资源不是孤立的, 相互之间存在着多种关联, 例如网页之间的链接、文献之间的引用, 不同个体的属性也会存在某种关系。图 4 中, r 的作者 u 属于某机构, 特定的实体对象, 如书、作者和机构通过属性建立了关系。对属性进行关联就能形成知识图谱, 从而为知识服务提供支持。综上所述, 信息组织乃至知识组织的一类核心工作就是在资源集合、个体、属性及其关系等不同层面定义操作。当这些操作可以被转化为自动实现方法时, 信息组织的效率将大大提高。

3 概念划分、分类体系及词间关系的数学解释

3.1 概念与集合

信息组织将概念逻辑原理作为一个基础理论, 编制分类法和叙词表都建立在概念划分的基础上。概念是逻辑学中的重要术语, 有内涵和外延两个方面。内涵是指概念所反映的客观对象的本质属性, 它回答对象“是什么”或“什么样”的问题。外延是指概念所反映的、具有同一内涵的客观对象的范围, 从量的方面去揭示客观事物, 它回答对象“是哪些”或“有哪些”的问题。从表 3 可以看出, 当概念内涵的限定性描述越多, 外延就越缩小。一些资料将两者的关系概括为“内涵越多的概念, 其外延越小”。需要解释的是, 这里“多”不是指范畴大, 而是指客观对象的本质属性多, 其表现为需要满足的限定多。

表 3 概念的内涵外延示例

概念	内涵	外延
三角形	指三条边首尾相连形成的多边形。	
直角三角形	指三条边首尾相连形成的、有一个直角的三角形。	
等腰直角三角形	指三条边首尾相连形成的、有两条边相等且有一个直角的三角形。	

逻辑学所谓的概念的关系, 实际是概念外延上的关系, 认为概念有相容和不相容之分。相容是指两个概念的外延至少有一部分重合, 不相容是指外延不重合。相容关系分为完全包含、部分包含和相等三种。完全包含中, 外延较大的概念被称为属概念, 较小的为种概念。从集合论的角度看, 外延的存在使概念可以看作是一组实例的集合。如图 5 所示, 每个概念就是一个集合, 概念外延的三种关系分别对应集合的包含、相交和相等三种运算。集合的运算有交、并、补、对称差等类型, 还遵循幂等律、结合律、矛盾律、排中律等一系列运算律, 这使得概念间各种关系有了进一步丰富的可能性。集合基本概念及运算可参阅集合论相关内容。

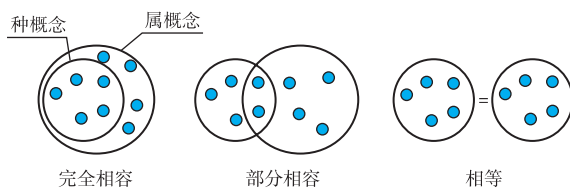


图5 用集合表述的概念外延关系

3.2 概念的划分与等价关系

信息组织用逻辑学的“概念划分”来表述对事物的分类。分类相当于对外延的划分，即在保持实例具有共同本质属性的前提下，按照非本质属性将实例划归为不同的子集，例如，三角形包括锐角三角形、钝角三角形等；图书包括纸质图书、电子书等。信息组织的分类原则规定“划分所得各子项的外延必须互相排斥，所有子项的外延之和应当等于母项的外延”；此外，还根据操作经验规定“每次划分必须按相同的标准进行”。事实上，集合的划分有精确数学定义对应上述文字的含义，而这段文字只说明了划分结果应具有的特征，却没有说明怎样得到划分。为了帮助理解，本小节简要引入集合论中的相关内容。

设两个非空集合，记为 C 、 D ， $C \times D = \{ \langle x, y \rangle \mid x \in C \wedge y \in D \}$ 为 C 、 D 的笛卡尔积，其任意非空子集 F 表示一种建立在 C 、 D 上的二元关系，如果 $C=D$ ，也称 F 是集合 C 上的二元关系。

对 C 上的二元关系 F ，可以考察其是否具有自反、对称、传递三种性质，如果兼备，则称 F 是等价关系。其中三种性质的定义如下：

若 $\forall x (x \in C \rightarrow \langle x, x \rangle \in F)$ ，则 F 满足自反性，例如实数集合上的小于等于关系就满足自反性，而小于关系不满足自反性；

若 $\forall x \forall y (x, y \in C \wedge \langle x, y \rangle \in F \rightarrow \langle y, x \rangle \in F)$ ，则 F 为 C 上的对称关系，例如人类集合上的兄弟关系满足对称性；

若 $\forall x \forall y \forall z (x, y, z \in C \wedge \langle x, y \rangle \in F \wedge \langle y, z \rangle \in F \rightarrow \langle x, z \rangle \in F)$ ，则 F 是 C 上的传递关系，例如实数集上的小于关系满足传递性。

如果某种二元关系 F 是等价关系，就可以依照集合 C 上哪些元素满足 F 所定义的等价关系，将 C 分成若干完全不相交的子集，且所有子集的并集恰为 C 。每个子集称为一个等价类，每个等价类内部的元素相互具有等价性。这样的特点完全对应本小节第一段中转述的信息组织分类原则。也就是说，分类体系依据概念进行类别划分，本质上是外延的集合划分问题。

例如，在集合 $C=\{1,2,3,4,5,6,7,8\}$ 上按照“除以3的余数相等”定义的二元关系 F 就是一种等价关系。由于 C 中的元素除以3的余数共有三种情况，所以产生的不同的等价类有3个，即 $\{1,4,7\}$ 、 $\{2,5,8\}$ 、 $\{3,6\}$ ，如图6所示。它们互不相交，但内部紧密关联。图中带箭头的连边表明起止两节点对应的元素满足关系 F 。当将 F 的定义改为“除以2余数相等”，那么 C 产生的划分有两个不同的等价类，即 $\{2,4,6,8\}$ 和 $\{1,3,5,7\}$ 。等价关系所满足的自反、对称、传递三种性质保证处于每个子集中的元素具有相同的本质，类似于语言中含义相近的词、或者某种场景中能互相替代的事物。



陈翀, 张婧莹, 李春秋. 计算思维视角下信息组织核心概念与描述模型的再认识 [J]. 文献与数据学报, 2021, 3 (4): 012-025.

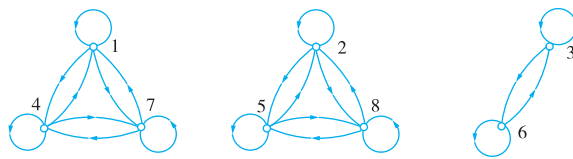


图 6 集合 C 上等价关系 F 产生的划分结果

具体分析信息组织所采用的学科分类体系或知识分类体系, 其中每个类别的子类都是由相应实例集合上的某种划分决定。产生划分的等价关系有些是按照年代、地域等简单量定义, 有些是依据学科的复杂内涵来定义。如文学可以按“年代”分为“古代文学”“现代文学”和“当代文学”三个子类; 情报学按现有的研究领域可以分为“情报检索”“情报资料处理”等多个子类。在实际中, 对概念内涵的理解和表述有语言模糊性, 所以事物的集合和集合上的关系不像数学量那么精确。在遵循等价关系基本性质的前提下, 人们会根据应用需要和对领域知识的习惯认识调整类别。

3.3 概念划分的数学描述

表 4 给出了划分的数学定义和信息组织对划分规则的对照描述, 二者本质一致, 表明可以从集合划分的角度理解并表述逻辑学中的概念划分。将概念外延所对应的实例集合记为 C, 只考虑 C 为非空集合的情况。若 C 的子集族 π 满足表 4 中的三个条件, 则 π 是 C 的划分, 称 π 中的元素为 C 的划分块。

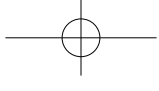
表 4 概念划分的数学描述

划分的条件	含义	信息组织中对划分规则的描述
$\emptyset \notin \pi$	划分的结果不能包含空集, 需要有至少一个元素。	母项析出的概念中不能有不存在概念。
$\forall x \forall y (x, y \in \pi \wedge x \neq y \rightarrow x \cap y = \emptyset)$	任意两个不相等的划分块无相交。	划分后的子项 (即划分出来的概念) 必须是互斥的。
$\cup \pi = C$	所有划分块的并集为 C 自身。	划分后的子项外延之和等于其母项 (即被划分的概念)。

3.4 叙词表词间关系的数学解释

叙词的词间关系主要有等同、等级、相关关系, 这是从概念含义上进行命名的, 它们实际上都是二元关系。等同关系中的元素具有相同的本质, 对应着等价关系中同一个划分块的元素, 这些元素构成的集合是等价类, 可依据其本质特征进行相互替代。

等级关系中的元素主要有属种或整体部分之分, 它们实际上是满足某种偏序关系。偏序关系与等价关系在性质上有所区别, 是指非空集合上具有自反、反对称和传递的二元关系。实数集上的小于等于关系、幂集上的包含关系都是典型的偏序关系。偏序关系导出了元素之间的可比性, 可比并非指数值大小, 而是按照偏序关系的实际含义决定元素之间的次序, 这是产生类目层次的基础。



相关关系根据事物之间的联系来定义,相关有多种多样,有些是因果所致,有些是事件或性质所致。这实际上对应着集合上的二元关系,表达了事物之间广泛存在的联系。集合中的关系是笛卡尔积的子集,并不一定有对应的表达式或实际含义,但在信息组织的类别体系或叙词表中,相关关系一定是具备某种实际意义的。

4 信息组织的逻辑结构

信息组织中常用目录和索引反映资源类别或个体的序关系,并形成一定的结构,利用离散数学的集合论和图论知识可以概括上述关系及结构。此外,考虑到查找资源时的搜索效率也是信息组织的一个重要目的,本文还借鉴数据结构组织数据的思路来分析信息组织逻辑结构的类型和适用情景。

信息组织中一般是通过比较资源的属性值来进行分类或排序。属性值是数据,因此在讨论信息组织的逻辑结构时,可以参照数据的组织结构。组织数据的有序结构一般分为线性、树状和网状三种。

线性结构是指集合中存在唯一一个“头”和唯一一个“尾”,除了头元素之外,每个元素有唯一一个前驱元素,除了尾元素之外,每个元素有唯一一个后继元素。这个结构如同一个按某种规则排好的队。

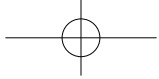
树状结构是一个连通无回路的图。通俗而言,一棵树有唯一一个头,除了头元素之外,每个元素有唯一一个前驱和零到多个后继。多棵树可以并存,构成森林。树结构如同一串葡萄。

网状结构是每个元素的前驱和后继数量都不限,元素之间的关系是多对多的。这个结构如同城市公交站点连接图。公交站之间由公交线路连接,一个公交站可以被连到至少一个其他站点。

由于树状结构符合人们对事物归属关系的理解,能表达的意思比线性结构丰富,查找其中的元素也比网状结构容易,现实中被广泛使用。但是它的数学定义要求一个元素的前驱不能超过一个,这对实际应用来说过于严格。因为资源的属性很多,即可分类的维度多,在一个大的类别体系中,某个资源属于多个类目或者一个子类兼有多个父类都很常见。如果放宽定义,在保持树的层级性同时也允许一个元素有多个前驱,就成为层状结构。因此,在信息组织中逻辑结构一般为线性、层状和网状三种。

层状结构遇到的问题是,要判断每个元素所属的前驱类别,判断代价大,而且结构欠缺灵活变化的适应性。特别是在新资源、新类目不断产生的情况下,在现有的层状体系中,有时无法找到对应的前驱,这对处于不断发展的资源集合而言很不方便。一种更为灵活的结构是网状。网络有节点和边,待组织的元素就是节点,元素之间的各式各样的关系就是节点间的边。新增节点只要与已有节点存在特定的关系,就可以加入到这个网络中。知识图谱就是网状结构。

资源对象视为节点,资源之间的关系视为节点之间的边,这些点和边就形成一张图。信息组织任务就是依据该图建立节点集合的序。具体采用怎样的逻辑结构取决于资源集合的特点和组织的目的。如果是侧重保存,则要考虑支持程序快速查找,可以选择线性结构,例如按资源某个属性的值进行排序,将所有资源组织成线性序列。如果是侧重服务,那要分析外界与资源的交互



陈翀, 张婧莹, 李春秋. 计算思维视角下信息组织核心概念与描述模型的再认识 [J]. 文献与数据学报, 2021, 3 (4): 012-025.

方式, 可以提供层状类目来帮助用户了解资源集合的概貌, 还可以从多个维度描述资源并建立网状关系, 让用户能以多种方式找到目标。表 5 归纳了三种结构的特点及适用情况。在具体的应用中, 三种组织结构可以被同时使用, 来解决用户搜寻信息的不同阶段中的需要。例如网页信息组织中, 层状的内容导航目录、线性的网页索引、网状的原始网页组织结构都是需要的。

表 5 信息组织的逻辑结构特点与应用实例

结构	不足	优点	实例
线性	将对象的特征简化到一维, 割裂了多种可能的关联, 可读性差。	容易编码, 查找效率高。	字典序、文档索引
层状	层状约束严格, 两个节点间可能产生交叉, 一个节点也可以有多个父亲。	反映了层次和类别的关联, 可读性好。	中图分类法
网状	结构发散, 查找不便。	结构有良好的增长性, 适于不断变化对象集合的组织。	知识图谱

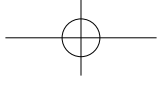
5 讨论

在应用需求的驱动下, 信息组织主要处理对象从传统文献扩展到网络资源、数据集甚至商品、人类等实体, 数据的种类和数量增加、关系更加复杂。这一过程中, 支持处理大规模复杂对象各类技术方法被引入信息组织的知识体系, 越来越多的词表工具或组织规范是与这些方法相适配的。经典组织理论诸如分类体系构建原则等, 多数是在传统对象的组织实践中获得的定性总结; 而基于自然语言处理和机器学习的自动分类、类目生成, 或基于语义网相关规范的本体描述等任务用到更多的是来自计算机领域的技术方法。如果能将信息组织发展演变中的知识形成一致的表述, 无疑能加强整个知识体系的紧密性。事实上, 在语义网背景下兴起的知识组织, 通过借助 RDF、OWL、SKOS 等典型的形式化方法描述资源和知识, 已经证明了这样的表述有利于系统间交互和资源共享。

此外, 信息组织在图情专业的课程体系中有重要地位, 课程内容主要围绕信息的描述、标引、分类等问题介绍相关理论、方法、标准规范及应用。信息组织中既体现着对组织对象的管理理念, 也依赖不同自动化程度的具体实现。现有课程内容的较大比例是来自前人对文献组织的经验归纳, 且与引用的支撑理论缺少密切联系。在专业学习中也因为缺少逻辑学、语言学、系统论等先修课程, 学生会对信息组织理论知识感到困惑。而计算机数学基础、编程基础的课程与信管、图情类专业课程中的关联是比较密切的, 将计算思维与信息组织理念相融合, 更有利于基础知识贯通。

6 结论

大数据时代是信息时代的延伸, 它不但具有信息时代的基本特征, 还对人类的数据驾驭能力



提出了新的挑战,这促使我们对信息组织的发展做出新的思考。本文的结论主要有三个方面:一是归纳现有的信息组织核心概念,用面向对象的方法抽象,不但统一了信息组织描述框架,还便于自动化系统理解和实现组织活动。二是借助离散数学中的相关概念,对信息组织中的概念划分、分类体系及词间关系进行形式化概括。与现有的信息组织知识表述相比,内容上相吻合,表述上更精练。三是借鉴数据组织和图论来归纳信息组织的逻辑结构,并分析了不同结构的适用情况。

信息总量的剧增,资源形式的差异,使得信息组织必须走实用化、自动化路线。资源往往关联着多维度数据,用户对资源需求也更加灵活多样。未来可以探索将资源描述、存储、分类、索引、访问等各类组织操作围绕资源对象进行定义,并融合计算思维将一些蕴含了数学及工程化思想的信息组织知识进行关联,逐步完善信息组织的支撑理论体系,使之在继承经典组织理念的基础上,凸显管理学与计算机科学交叉的特色,为信息组织的理论实践更好地与自动化技术相结合提供概念表述体系上的依据。

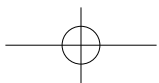
【注释】

① <http://upload.wikimedia.org/wikipedia/en/3/37/Semantic-web-stack.png>.

② <http://www.dlib.org/dlib/July95/07arms.html>, 1995, d-lib magazine by William Y. Arms. (数字对象模型的提出源于: <http://www.cnri.reston.va.us/home/cstr/arch/k-w.html>, cnri.dlib/tn95-01, 1995, by KAHN R, WILENSKY R)。

【参考文献】

- [1] 赵一鸣, 马费成. 大数据环境对信息组织的影响 [J]. 图书情报知识, 2017(1):4-10.
- [2] 罗龙艳, 杜茂康, 张仿, 等. 信管专业的信息组织、存储与检索课程改革初探 [J]. 情报探索, 2008(1): 25-28.
- [3] 中国大百科全书出版社. 中国大百科全书第三版网络版: 信息组织 [EB/OL]. [2021-11-01] <https://www.zgbk.com/ecph/words?SiteID=1&ID=59072&Type=bkzyb&SubID=60801>.
- [4] 魏敏. 信息组织 4.0: 变革历程和未来图景 [J]. 国家图书馆学刊, 2018, 27(1):78-85.
- [5] 张敦仲. 信息组织学课程体系建设 [J]. 中华医学图书情报杂志, 2009, 18(2):74-77.
- [6] 贾君枝. 面向数据网络的信息组织演变发展 [J]. 中国图书馆学报, 2019, 45(5):51-60.
- [7] WING J M. Computational Thinking [J]. Communications of the ACM, 2006, 49(3):33-35.
- [8] KAHN R E, WILENSKY R. A framework for distributed digital object services [J]. International Journal on Digital Libraries, 2006, 6(2):115-123.
- [9] 马张华. 信息组织 [M]. 第3版. 北京: 清华大学出版社, 2008.
- [10] 周宁. 信息组织 [M]. 第4版. 武汉: 武汉大学出版社, 2017.
- [11] 戴维民. 信息组织 [M]. 第3版. 北京: 高等教育出版社, 2014.
- [12] 叶继元. 信息组织 [M]. 第2版. 北京: 电子工业出版社, 2015.
- [13] [美] 泰勒. 信息组织 [M]. 张素芳, 李书宁, 李金波, 译. 2版. 北京: 机械工业出版社, 2004.





陈翀, 张婧莹, 李春秋. 计算思维视角下信息组织核心概念与描述模型的再认识 [J]. 文献与数据学报, 2021, 3 (4): 012-025.

Further Understanding on the Core Concepts and the Description Model of Information Organization from the Perspective of Computational Thinking

CHEN Chong ZHANG Jingying LI Chunqiu

(School of Government, Beijing Normal University, Beijing 100875, China)

Abstract: [**Purpose/significance**] In the era of big data, the information is complex and of large scales, which brings an inevitable challenge for information organizations. In order to adapt to smoothly transform the human knowledge of information organization into the computer understandable form, this article intends to extract the mathematical and engineering ideas from the theoretical knowledge of information organization, defines the core concepts in a formal way, and helps to improve the scientificity of knowledge representation in information organizations. [**Method/process**] Selecting core concepts of information organization, this study aims to find the knowledge of similar essence with these core concepts from discrete mathematics and object-oriented modeling. Based on computational thinking, the similarities are compared between the core concepts of object-oriented modeling and the information organization, such as classification and metadata definition. The content of concept division and the construction of classification system are explained from the perspective of set division. The term relationships in thesaurus are expressed with binary relation theory, and the logical structure of information organization is summarized by graph theory. [**Result/conclusion**] This study establishes an object-oriented information organization description model, reveals the mathematical connotation of concept division, classification system construction and term relationship in thesaurus; and also analyses the characteristics of three logical organization structures. The innovation of this study reflected in: it makes the qualitative knowledge of information organization to be abstract with computational thinking, and promotes the formal representation of core concepts. It not only provides a basis for using the properties of relevant mathematical concepts to improve the knowledge system of information organization, but also explores the better integration between traditional qualitative theory and modern technology in information organization, which helps consolidate the consistency of the theoretical system of information organization.

Keywords: Information organization; Computational thinking; Object-oriented; Descriptive model; Term relationships; Construction of classification system; Logical structure of information organization

(本文责编: 孔青青 孙龙慧)