

生物医学知识整合论(Ⅳ)
——关于知识的深层次哲学思考

包含飞

上海中医药大学中医学信息化-标准化研究室, 200032

摘要 本文继续讨论与生物医学知识整合或一般知识工程密切相关的人类认知科学的深层次问题。讨论了信息的抽象性和具体性及医学信息与香农描述的信息的区别,并指出医学信息的具体化、个性化和较少一般性等特点来自生物医学系统纷繁复杂的结构。接着文章对与知识表示、处理、操作或整合密切相关的32个哲学基本概念加以定义和描述。最后分析了各种知识表示形式的着重关系和着重结构的两种主要倾向。

关键词 生物医学知识整合论 生物医学信息学 医学人工智能 知识表示

(I). 香农的抽象信息与生物医学的具体信息

无论是德国科学家鲁道夫·克劳修斯和奥地利物理学家波尔兹曼提出的熵,还是美国数学家香农描述的信息,都与哲学意义下的事物的某种均匀性或不均匀性的量度有关。是十分抽象和一般化的概念。

然而生物医学信息是程度不等的具体化信息,是形形色色生物医学结构规定下的信息。由JH van Bemmelen和MA Musen主编的《医学信息学》给出了如下医学认知过程:数据—(解释)→信息—(归纳)→知识。该书指出:“临床医生在观察病人或某一(生物的)过程中产生数据,并从这些数据中通过解释或推理得到信息……”,“通过仔细地研究大量医学方面诸如此类的解释过程,或者通过收集来自大量病人的数据解释,最后归纳推理得到新的见解和新的知识。”^[1]这里,我们看到的《医学信息学》给出的信息是夹在生物医学具体数据和具体知识之间的信息,因而它本身必定是具体的。以笔者的理解,数据、信息及知识有如下关系:数据:原始观测的记录—→信息:数据之间的关系—→知识:数据之间的更深层的关系(信息和知识之间并无明确的界限)。

作为具体的信息,与香农等提出的一般意义下的极度抽象的、极其基础的信息概念有天壤之别。我们看到无论是数据上升为信息,还是信息上升到为知识,都必须经过“解释”或“归纳”以获得新的层次或性质不同的关系(新的层次或性质不同的语义)。问题就出在这里,较之其他领域,生物医学对象的一般性(统一性)明显较少,个性(差异性)明显较多,因此这种“解释”、“归纳”及由此而获得的“新的语义”也是具体的,明显地较少一般性,带有丰富多采的个性或差异性。

以生物医学的“基础理论”为例说明一般性和具体性在该领域中的尖锐冲突:与其他科学比较,生物医学各个领域的“基础理论”如DNA指导蛋白质合成的中心法则,酶化学理论,细胞分化理论,神经递质和调节理论,第二信使理论,激素-体液调节理论,体液和细胞免疫理论,载体理论,受体理论,离子通道理论等等,很多情况下仅仅是生命哲学层次的一些理解,其操作性和临床指导意义(如细胞分化理论对癌症治疗的指导性,神经分布对逻辑思维的解释性)十分有限。生命系统是一个复杂的同质-异质的、同构-异构的或复合结构的世界。基础理论的一般性、普适性、应用性和解释能力均受到无情的挑战。就象“在崇山峻岭上耕作”拖拉机一样举步维艰。经验、实验和现象还是“绝对权威”,不容你不承认,

科学理论常常不得不低下“傲慢的头”，向复杂多变的个性折腰致敬！医学还是经验论者和手工操作者的天堂。一个责任心强的好医生在其工作中必须永远睁大双眼，边判断-边观察-边处理。这就是医学信息世界！

引起生物医学信息的个性和差异性的根源是生物医学的无与伦比的丰富的结构。因此任何生物医学中的事物，除了考虑它们的独立形态或本体形态（如在科学实验中把它们分离后的形态）外，还必须考虑它们的非独立形态或附体形态。二者对应于事物的两种不同状态，前者相对于事物的独立态、自在态、个体态、惯性运动态、全潜能态。为哲学本体论追求的目标；哲学的本体有事物的全体、真体之意（也许它是个无穷概念），是较为理想化的概念；而后者对应于非独立态、组合态、应用态、环境约束态、联合态、限定功能态。此为应用领域本体（domain ontology）追求的目标。应用领域本体研究为应用目标所驱动，只研究事物的相关部分，因此是较为物理性的概念。实际世界的事物多处于附体形态，处于形形色色的结构之中，因此同一事物的物理属性各不相同。

通过科学手段获得知识的过程，往往从分解对象（还原）开始，而分解对象就是分解结构，破坏整合。在我们希望获得这些元件的本体时，我们牺牲了整体事物的本体。就象我们希望了解水（ H_2O ）的本体，如果我们把它分解为氢（ H_2 ）和氧（ O_2 ），那么即使我们完全了解了氢和氧的本体，我们仍无法了解水的本体。从独立意义下事物的本体到该事物作为其他事物的组件或附体有一个属性选择过程。如为扑灭房子失火，我们站成一个队列传递水桶救火，此时每个人都处于非独立态中，除了作为传水链上的一环外，万能的人类别无他用。这就是人作为这一结构的附体所表现的属性。

任何把生命系统简单化的企图注定会碰得头破血流。如果把如此纷繁复杂的充满各种结构语义的医学信息理解为哲学或一般意义下的信息，把生物医学系统回归到“扔骰子”系统，那么医学信息学势必难于深入到生物医学的核心中去，只能停留在充当“工具”或“计算机在医学中的应用”的角色。这里，可以对照一下心电图、B型超声波、X-线、CT等在医学中的成功，由于这些新技术创造了很多新型的生物医学数据和内容或语义，所以在谈论它们时，人们的注意力自然就集中在它们所创造的新型的医学内容或语义，没有人会在乎心电图、B型超声波、X-线和CT的机器设备及其软件究竟是什么。这给我们带来一个重要启示，毋庸置疑医学信息学作为无与伦比的新技术定能开创无与伦比的医学数据和语义的新世界，但这一天尚未到来。欲穷万里目，更上千层楼，但登楼所付出的代价将大得多。医学信息学家必须经过比其他领域更为艰苦卓绝的努力，才会迎来他们值得骄傲和庆祝的日子。

（II）. 概念和定义

知识可看作是客观物理世界的在我们意识世界中的模型。反之，客观物理世界是知识和意识世界的原型或母型。从根本上讲，知识和意识依赖于客观物理世界而存在；但在相当程度上，知识和意识世界具有自己的独立于客观物理世界的运动和发展规律^[10]。而作为“第三者”，数据-信息-知识系统是物理世界与意识世界二者模型的混合，既有物理性的一面，又有意识性的一面。这是我们发展知识工程包括知识整合工程^[2-12]首先要明确的基本哲学关系（见图1）。

我们不可能离开知识和它们所对应的物理或逻辑对象的丰富多彩的哲学属性来对知识进行表示、处理、操作或整合。因此讨论下述与知识密切相关的基本概念是必要的。

定义1 知识的背景空间：为知识表示形式以外的、潜在的影响或决定其意义的其他事物的总称，即知识的意义赖以成立的条件。包括广义的知识参量维度或参量空间；欧几里得几何的5大公理组成的逻辑空间；牛顿力学赖以成立的时间、空间的均匀性和质量的恒定性

等形成的哲学空间及爱因斯坦的相对论所需的光速恒定-时空不均匀-质量速度互变的哲学空间；也包括表达知识的主体的视点、视场、意向、意场（如“猫吃老鼠”这一事实，对猫来讲是一顿美餐，对鼠来讲是灭顶之灾，对生态学来讲家是物种的平衡行为）；还包括我们的所谓标准知识或公认知识（包括本体学）赖以成立的我们至今还没有作过详细探讨的人们公认的逻辑空间（包括人类感觉系统所依赖的背景空间，其中时间、空间的均匀性和质量的恒定性是其核心成分）等。

也许由于大脑固有的功能特征所致，这些背景空间往往潜伏于我们的意识之下，致使以往的一些逻辑科学或人工智能研究误以为我们的知识及其运动规律是无背景的“裸露”的对象，待到这些系统被应用于背景复杂的实际（特别是生物医学实践）中时，就成了“弱不禁风的林黛玉”。

知识背景空间应该是有限空间，不是无限空间，即应该是可操作的。]

而生物医学知识系统的背景空间是多变的，不统一的。多变的知识背景空间创造了生物医学知识的复杂性，是对生物医学知识整合的一大挑战。

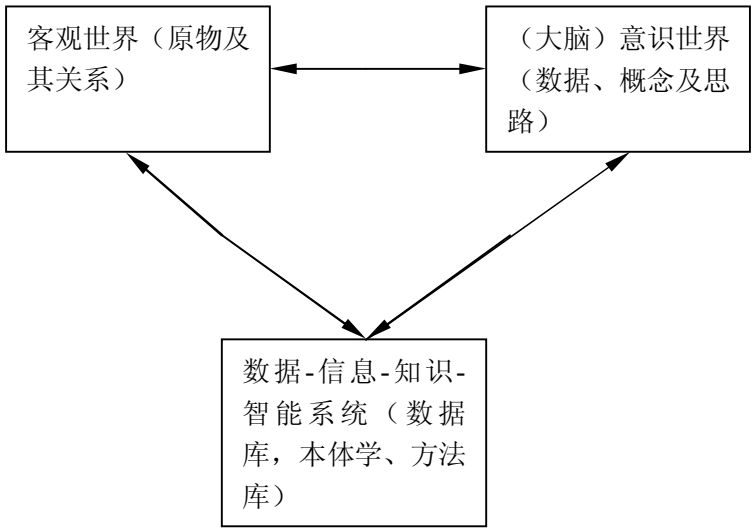


图 1 原物世界，意识世界与基于知识的智能系统的关系。

定义 2 广义参量维度或参量空间：知识所依附的背景空间之一，即知识生成所依赖的观测、干涉方法和维度的整合。涉及人们是通过何种（广义和狭义的）观测器（维度）、何种观察粒度来观测和表达事物，产生数据、信息和知识，用肉眼、墨镜、望远镜、显微镜还是某种仪器？

如火车上的旅客，对车站来讲他们是运动变的，对火车来讲他们是静止的。

维度是组成空间的元素，可视为一种惯性运动的可能性，有时可视为一种综合量度。

定义 3 概念：“定义一类对象的综合观念。”^[1]更具体讲：在某一知识背景空间中按其属性或行为可视为整体的单个事物或由某些关系联系起来的一组事物在意识中的对应体。概念对知识背景空间有依赖性，不是“赤裸地”存在的。

目前只有大脑具有创造概念的能力，计算机还不具备概念创造能力。

定义 4 有意义知识背景空间：与知识操作主体的（物理的或意识的）目标有关的知识

背景空间。有意义知识背景空间以外的任何空间为无意义知识背景空间。

例如拳击台上，观测维度或胜负标准是拳手获得的点数及是否能击倒对方，也包括一系列如拳手的体重级别、动作的规范、点数或违规的判法等组成的广义的有意义知识背景空间。其他如拳手的学历、相貌、婚姻状态、人种等组成广义无意义空间。

定义 5 结构：指事物在某知识背景空间中某种相对固定的、有意义的非均匀态或一组（物理的、意识的）关系的复合及其形式，反映了一组关系的不可分性捆绑性。狭义的结构指以笛卡儿三维空间为背景空间的排列、比邻或组织形式。广义的结构背景空间是由任何维度（物理的、意识的）组成的广义空间。

关系为元素之间的相互作用，两个以上的关系复合为结构。因而结构为三元以上的事物。著名的三体问题是个结构问题，知识整合是个结构问题，框架和形象是一种结构。

定义 6 结构形成机制：事物结构的形成的动力学机制，包括自组织机制，他组织机制，自-他组织机制。

定义 7 自组织机制：开放系统本身的一种内在信息积累或结构形成机制。如生命的起源、细胞的起源、物种的分化、植物、动物的生长发育。一般来讲，对于物理性自组织机制，人们的意志只能从外部进行有效干涉，如对输入或输出加以干预。如对庄稼进行施肥，拔除杂草等。在医疗实践中，还可以通过克服异体的排斥性，进行机体部分“移植”或“零件替换”。

定义 8 他组织机制：由（意识的、物理的、人为的、自然的）外力形成结构的机制。包括各种人工产品如各种机器的制造与组装的机制，其他物理或逻辑模型的建立的建立机制。

定义 9 纯意识他组织机制：意识作为外力形成结构的机制，故为一种特殊的他组织机制。包括数学概念如平行四边形、等腰三角形、等边三角形、圆等的形成。

定义 10 自-他组织机制：外力和内在力共同完成结构形成的机制，包括意识指导下的物理性自组织过程。

定义 11 意识-自组织机制：意识作为外力施加于自组织机制的结构形成机制，故为一种特殊的自-他组织机制，如基因工程、多利羊实验等。

定义 12 人工物理结构：为他组织机制形成的一种结构，即可用人工物理零件组装的结构，为概念系统指导下建立起来的物理结构，如钟表、飞机、火车、计算机和其他工业产品。

定义 13 人工意识结构：为纯意识他组织机制形成的一种结构，由概念、概念关系和概念运动组成的结构。包括各种知识表示方法及其推理模式、软件系统、数学算法、抽象的问题解决方法等。

例如“张三给李四一本书”这一生活中常见授受行为就可生成一个逻辑结构。如用语义网络表示则为^[13]（见图 2，图 3）：

讲到底，所有基本逻辑结构均源自基本物理现象及结构。

定义 14 自然物理结构：为自组织机制形成的一种结构，非人工零件组装的一种物理结构，由自然规律（或自组织机制）指导下建立起来的结构。如植物、动物等生物结构。

定义 15 人工-自然物理结构：如意识-自组织机制形成的一种结构，在人工环境下由自组织规律机制（自-他组织机制）建立起来的结构。如各种组织工程器官，基因物种，器官移植或嫁接等。

定义 16 人类意志可控性：人类意志对事物的干预程度或干预能力。包括全可控性、部分可控性和不可控性。事物结构形成中人类意识参与程度为人类意志对该结构的可控性大小标志之一。如果事物是纯意识构成的或意识-物理构成的，那么人类对它是全可控的或部分可控的。人类意志干涉包括输入、输出干涉及全结构或部分结构干涉：如零件替换，全组装等。

定义 17 结构维：由某种结构机制的广义的惯性运动形成的维度。如分形结构性。

定义 18 原物：客观世界万物。

定义 19 状态：事物属性、关系或运动的形态。可分为动态和静态。

定义 20 静态：事物的状态对其参量维度或参量空间的有效域（或意义域）来讲是个常态，不变态。

定义 21 动态：事物的状态对其参量维度或参量空间的有效域（或意义域）来讲是个变化态。也可能是一种序关系（如因果关系序）的整合形式。该序的一系列的事件的稳定性是如此之小，对于某一参量维或参量空间（如人类的感觉空间）的观测粒度来讲已融合为一个连续的动作（movement），无法辨别其各个定态。如电影，动画片等。动态事物的知识、知识表达、知识运动的规律与静态事物有很大不同，叫机器人搬移一个凳子和捉住一只兔子其难度不能同日而语。因为兔子的位置对参量空间笛卡儿空间来讲是不确定的，而机器人的操作需笛卡儿空间中的确定位置。

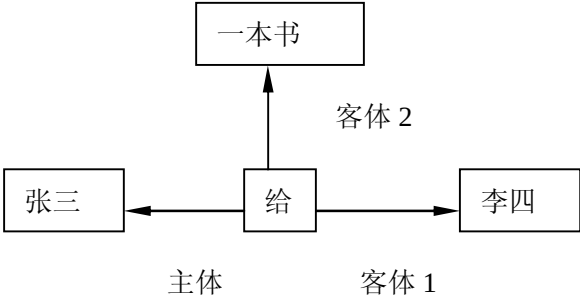


图2 带有动作结点的语义网络

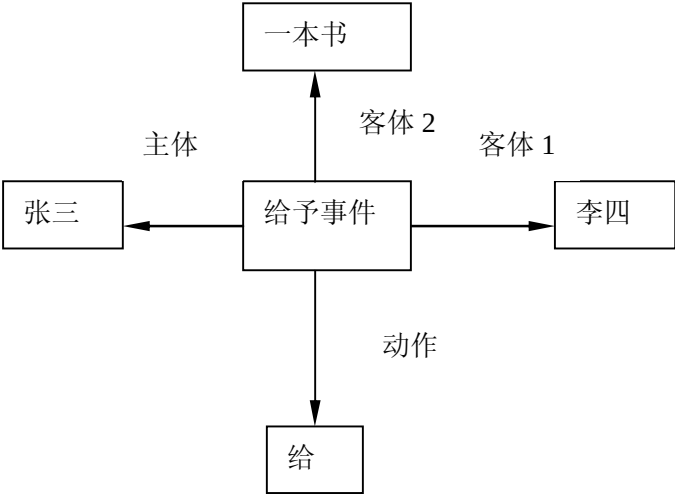


图3 带有事件结点的语义网络

定义 22 事物的一般属性集（本体学）：指事物（在人们公认的或默认的背景空间中）

作为一个独立实体意义下的属性集，也称理论属性集或概念属性集。理论上，只要对事物的观测参量的类型是无穷的，那么一般属性集的元素也是无穷的。

定义 23 事物的结构属性集（附体学）：某事物在非孤立情况下即作为元素参与某结构时所表现的属性的集合，它是该元素的一般属性集合（或本体学）的元素或子集，也称实际属性集或物理属性集。

定义 24 结构的属性限定：某事物作为元素参与某结构时，该结构作为整体对其属性的表现性（外显性或激活性）的选择，被“选上”的属性为表现属性（显性或激活属性），其他属性为非表现属性（隐性属性或非激活属性）。

定义 25 物理硬性（或稳定性）：事物属性、状态、关系或运动在某参量空间中持续（或稳定）程度。如阿庆嫂的茶馆是物理硬性事物，而南来北往的客人是物理软性事物。“一辆奔驰的火车”在笛卡儿空间中是不稳定的，但如果其作匀速运动，那么把它在速度空间中是稳定的。物理硬性一种度量为事物属性、关系或运动状态的持续的时间。物理硬性的反性是物理软性。

与此对照的是逻辑的硬性。

定义 26 知识的物理硬性（或稳定性）：事物物理硬性或稳定性在其知识中的反映。知识的物理硬性的反性是知识的物理软性。

与此对照的是知识的逻辑的硬性。

定义 27 概念粒度：概念的外延元素（子概念或实例）数量的倒数。可用

$$v = 100 / \text{外延元素（子概念或实例）数量}$$

表示。专有名词所代表的概念的 v 为 100，淋巴细胞对于血常规检查 v 为 100。但如需了解病人临床免疫状态，因为淋巴细胞可以进而分为 T-淋巴细胞, B-淋巴细胞, K-淋巴细胞, NK-淋巴细胞；而 T-淋巴细胞可再分为辅助性 T-淋巴细胞，抑制性 T-淋巴细胞，细胞毒性 T-淋巴细胞；细胞毒性 T-淋巴细胞又因为所感受的外部的或内部的抗原种类繁多，按其对抗原的特异性反应又可分为很多群体。因而在这一意义下，淋巴细胞的粒度远远小于 100。

概念粒度可影响概念的可操作性。

定义 28 认知黑洞：如事物由量变到质变是个突变，人们往往无法精确地知识描述突变的“过渡地域”，此为认知黑洞之一。其他如混沌系统的初态敏感问题，也是无法描述和控制的“地域”，故为另一种认知黑洞。

定义 29 抽象化：从较具体的实例集合或类集合中抽取若干共同属性，按后者组成较一般的新类的过程。

定义 30 思维轨迹：思维运动的一种属性或动态模式，包括抽象的问题解决方法。对于一些重复频率较高的问题类型，人类相应的思维运动形态也表现出相对的稳定性，这种思维运动形态减去其具体对象剩下的框架被称为思维轨迹。又可分为固有的基本思维轨迹和各种抽象程度的面向问题的解决方法。前者包括基本逻辑模式，如普通逻辑学的三段论。

问题类型重复出现的频率决定其思维轨迹的普遍性大小。

定义 31 黑箱化：事物的结点化、质点化或无结构化。事物的局整基本问题之一，知识的一种质变。大脑思维运动在必要时会采取黑箱化、结点化、二端化、经济化的运动形式，隐去其内部结构，只留下输入输出。通过黑箱化原理，大脑完成一种认知的跳跃过程（飞越大大小小的认知黑洞）。

此过程可导致大脑思维背景空间大变换，因此黑箱化为一种认知的(量变到)质变。

定义 32 语义的尺度依赖性原理：在线性意义下，随着系统（广义）尺度的不断扩展，在小尺度意义下的元素、关系、机制、结构、语义的个体重要性不断地降低，一旦超越了它们的有效尺度(有意义空间或维度)，这些元素、关系、机制、结构、语义将失去其个性而成为“尘埃”。而由它们融合而成的复合体（总体）将成为元素且重要性不断扩大，最终仅仅

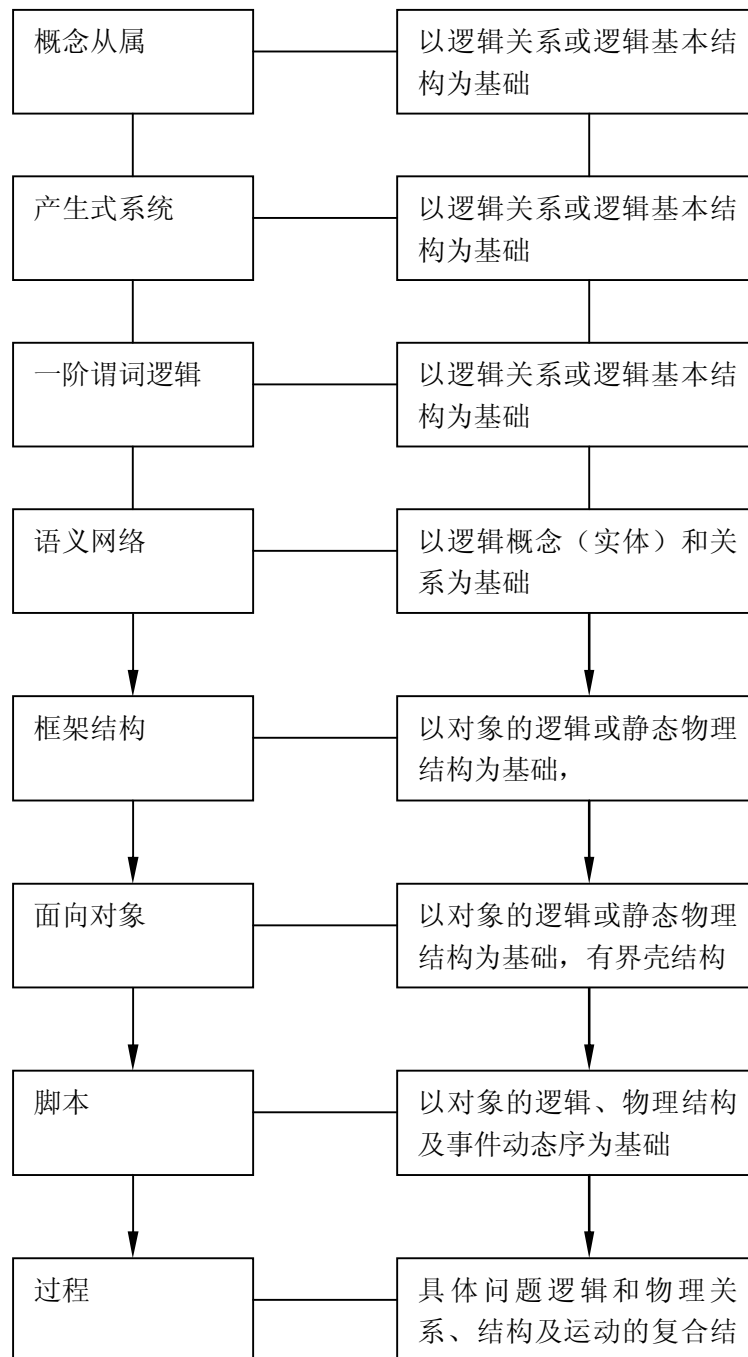


图 4 各种知识表示形式的意识性及物理性、关系性及结构性分析。前四者以关系和思维基本机制为基础，并不顾及对象的结构。从基于框架系统开始以面向描述对象的结构为基础。对象的所有属性以捆绑形式出现。面向对象知识表达是有“边界”的结构，禁止外界对内部结构的更动。而脚本作为特殊框架其“场景”槽描述一系列有序事件。过程则是关于具体问题的关系、结构及其的运动规律复合结构。

表现出总体的质。

例如随着宇宙范围的扩大，我们的地球及其“行星兄弟”的重要性越来越弱小，而由它们组成的太阳系，随之是银河系等新的复合体系的重要性越来越大，地球由“中心”，“小弟弟”，“无名之辈”，最后成为“灰尘”。

但诸如人体的微量元素等非线性机制不受此原理支配。

(III) 知识表示方法与事物的结构性之间的关系

知识表达方法大致可以分为两类：(1) 以大脑自身的逻辑或思维运动规律为基础，着重事物之间的关系及基本逻辑结构；(2) 以大脑所处理的意识或物理对象为基础，着重事物形形色色的结构。因为较为稳定的逻辑结构往往都有坚实的物理结构作为其背景（如银行信息系统有着较为稳定的银行物理运作过程为背景），所以后者主要与物理结构有关。物理结构及其不可分性决定了逻辑结构及其不可分性。知识表达从基于规则（关系）的产生式系统向基于框架的知识系统转化，是从着重关系向着重结构的转化，因为框架就是一种抽象的结构。

人工智能研究了很多知识表达方法，有关这些知识的具体形式、潜力和限制请读者阅读有关的人工智能及知识工程著作^[13-15]。本文主要从总体角度对知识的意识性和物理性、关系性和结构性加以讨论。概念从属、产生式系统、一阶谓词逻辑表达方法与语义网络虽然知识表示的具体结构有所区别，但它们均以关系或关系复合（简单的或基本的逻辑结构）为基础来表达知识和演绎知识运动，带有明显的逻辑主导性。而框架结构、脚本、面向对象等知识表示方法则强调对象具体的逻辑或物理结构，强调关系的捆绑性或不可分性，带有明显的物理主导性。

图 4 是关于这一观点的总结。

参考文献

1. JH van Bemmelen, MA Musen 主编，包含飞，郑学侃主译：《医学信息学》，上海：上海科技出版社，2002
2. Bao H.F.: The quantitative integration of biomedical information by computer--New research QMSOC(II), The Proceeding of the First Conference of the Frontier for Life Sciences in Central-south China, 1989, 216-222
3. Bao H.F., Geng J.H. and Su Z.F.: Pansystems Methodology(PM) and a new research on large-scale integration of biomedicine—An introduction of QMSOC and its recent progresses , Acta of Jiansu Industrial College, Journal of Jiangsu Institute of Technology, 1991, 4 (2): 69-75
4. Bao H.F., Ni X.W., Lou S.: Integratable Relationized Medical Electronic Book (IRMEB)--- An Exploration of An New Type of Intellegent Knowledge Medium Under the Influence of Pansystems Theory. Advances in Systems Science and Applications (Inauguration Issue) 1995,p304-309
5. Bao H.F.: HCSL:A Human-Computer Commonly Understandable and Communicatable Medical Language, 《Proceedings of The First China-Japan-Korea Joint Symposium on Medical Informatics(CJKMI'99)》, p177-181,1999
6. 包含飞：创造一种人机共解互通的医学语言(西医学部分)，第八届全国医药信息学大会论

文集, CMIA'99(电子工程师增刊), 1999, 98: 18-24

7. 刘沁, 包含飞: 人机共享语言 HCSL (医学) 的第二版的计算机实现, 中外名医杂志, (1): 78-80, 2001
8. 包含飞: 生物医学知识整合论(一), 《医学信息》杂志, 16 (6): 174-279, 2003
9. 包含飞: 生物医学知识整合论(二), 《医学信息》杂志, 16 (8): 410-415, 2003
10. 包含飞: 生物医学知识整合论(三), 《医学信息》杂志, 16 (待发表) 2003
11. Bao H.F.: The Theory of Biomedical Knowledge Integration(I), Chinese Journal of Medical Treatment, 2(13): 1-6, 2003
12. 包含飞: 生物医学知识整合科学思维的成分异源性 (I), 上海计算机用户协会年会论文集 (2003), (待发表)
13. 王万森: 《工智能原理和应用》, 北京: 电子工业出版社, 2000
14. 林尧瑞, 马少平编著: 《人工智能导论》, 北京: 清华大学出版社, 1989
15. 陈世福, 陈兆乾等编著: 《人工智能与知识工程》, 南京: 南京大学出版社, 1997, 109-116