

生物医学知识整合论(II)

包含飞

上海中医药大学(200032)

中医学信息化-标准化研究室

摘要 本文进一步阐明生物医学知识整合(BMKI)理论的意义,并提出生物医学知识整合的一种基本原则——求供覆盖原则,并以心脏泵血循环为例说明了这一原则。文章的第三部分以UMLS的语义网络为例初步讨论了知识的二重性即物理性和意识性,又讨论了它们对BMKI的可能影响。

关键词 生物医学知识整合 UMLS 生物医学信息学

在生物医学信息学领域中,我们可以看到大量令人钦佩的和催人奋进的关于医学数据、信息和知识的表达、存储、提取、处理和应用的研究和开发^[1]。笔者的信念是“浩如烟海的生物医学的数据、信息、知识及由它们组成的形形色色的学科分支既然来自于同一个故乡——不可分割的机体,它们最终都将“回归故里”,重新整合成为不可分割的(当然是虚拟的)机体。”因此生物医学知识整合(Biomedical Knowledge Integration, BMKI)^[2-12]是一切生物医学逻辑的自然归结,也是生物医学学科的大联合、大结合的一种探索,其最终目的是求出机体各部分(正常的、异常的、同质的、异质的)模型的总和。BMKI也是当前生物医学信息学研究热点“数字虚拟人体”工程^[13-16]的一种必不可少的基础研究。很难想象可以跳越BMKI这一基础研究来谈论任何类型的“数字虚拟人体”工程。讲到底,“数字虚拟人体”工程就是知识整合工程。无论是虚拟解剖人、虚拟生化人、虚拟药理人、虚拟生理人,还是虚拟心理人,乃至虚拟“天才人”,概莫能外。当然要达此目标,不可能离开对各种各样的异质的生物医学数据、信息和知识的信息学性质的艰苦深入的研究。

(一) 盲人摸象新说

一个为大家所熟悉的寓言说有五个盲人摸索一头大象,盲人一摸到大象的鼻子,就声称“大象如一根管子”,盲人二根据大象的尾巴,说“大象如一根绳子”,盲人三、四、五各自摸到大象腿、躯体和耳朵,就分别认为大象酷似“柱子”、“墙”、“扇子”。此说一直被人们作为讽刺以局代整、以偏盖全的典型笑料。但事物总是不断进步的,后来盲人们掌握了“整合理论”,他们不再坚持一己之见,而是用附属关系和空间分布关系来整合他们五人所获得的信息。整合的结果是:大象的“中心为一堵墙,墙的前面附着管子和扇子,下面附着柱子,墙的后面附着绳子”(见图1)。利用信息整合技术,我们的盲人小组获得了巨大进步。虽然他们各自所拥有的信息丝毫没有增加,但他们所构建的大象模型向真实的大象逼近了一大步。借此笔者希望阐明生物医学知识整合论的意义:虽然研究者的观测视点、观测方法、处理观测结果的技术千差万别,但所有的生物学、医学及生物医学信息学的研究和开发都是环绕着机体这同一目标,他们的结果之间必然存在内在的联系。深入探讨、挖掘、建立这些联系有助于我们在整体上对机体的理解,这就是BMKI的根本任务。

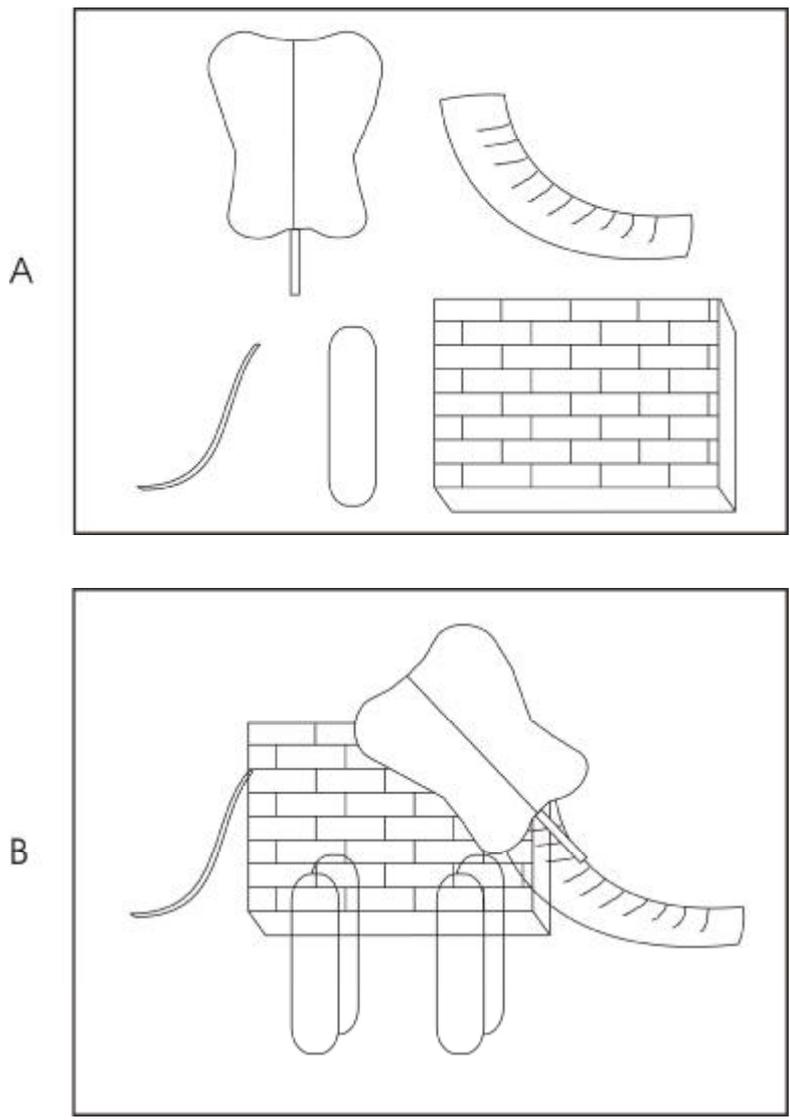


图 1 A 为在应用整合理论以前盲人们心目中的大象，
B 为在应用整合理论以后盲人们心目中的大象

(二) 求供覆盖——整合的一种基本原则

泛系理论^[17]指出了事物之间的供求索交原则，这是关系复合的一种基本机制。任何事物及其运动都可以解析为关系的复合。但任何关系的成立或实现都需要一定的或明或暗的条件或环境保证。为保证关系实现的可行性，这些条件或环境必须是充分的，为保证其经济性以减少资源浪费，它们又应该是必要的。任何关系均如此，无一例外。现在我们来谈一对关系序偶，如果前关系实现的结果可以充分满足后关系实现的条件，那么这种供求索交就称为供求覆盖。只要这对关系序偶是供求覆盖的，那么它们就可以复合。因为当我们讨论一串关系复合时总是不断地把中心向关系对的后关系转移，所以与“供”相比，“求”是问题出发点，故这一原则在本文中称为求供覆盖原则（Supply-Demand Covering Principle）。

由于机体运动的各种关系或机制的具体物理过程极为复杂，而且它们之中的绝大多数细节至今尚未阐明，对机体的各种物理机制象钟表一样了如指掌几乎不可能。所幸的是关系实现所需的充分条件往往是可

以通过实验或经验获知的。在一般时空条件和具体条件如物质和信息得到充分满足的情况下，关系就必然实现。至于我们是否了解关系实现过程的具体细节，从整合操作的角度来讲，是无关紧要的。因此考察两个关系是否能够复合或整合，只需考察前关系实现的结果（供集合）是否覆盖（或包含）后关系实现的前提（求集合），如果覆盖为真，则后关系必然实现。前后两个关系复合或整合即告成立。

求供覆盖是我们考察一个整合能否实现的基本而经济的原则，也是我们试图开发数字生理人体、数字生化人体、数字药理人体等虚拟人体的一个基本而经济的原则。

现在我们用求供覆盖原则来观察心脏泵血循环中多关系整合的实现。根据心脏泵血的生理过程^[18-19]，我们先设定若干具体的转化关系和转移关系，并引进若干激发心脏泵血循环必需的外部关系：

- (1) 半月瓣开启状态—[转化关系]—→半月瓣关闭状态；
- (2) 半月瓣关闭状态—[转化关系]—→半月瓣开启状态；
- (3) 房室瓣开启状态—[转化关系]—→房室瓣关闭状态；
- (4) 房室瓣关闭状态—[转化关系]—→房室瓣开启状态；
- (5) 心房内血液—[转移关系]—→心室内血液；
- (6) 心室内相对低压状态—[转化关系]—→心室内相对高压状态；
- (7) 心室内血液—[转移关系]—→主动脉内血液；
- (8) 心室内相对高压状态—[转化关系]—→心室内相对低压状态；
- (9) 外部关系：心房肌舒张，心室肌收缩，心室肌舒张。

那么基于求供覆盖原则我们就可以用图 2 描述心脏泵血循环中一系列定性整合过程。

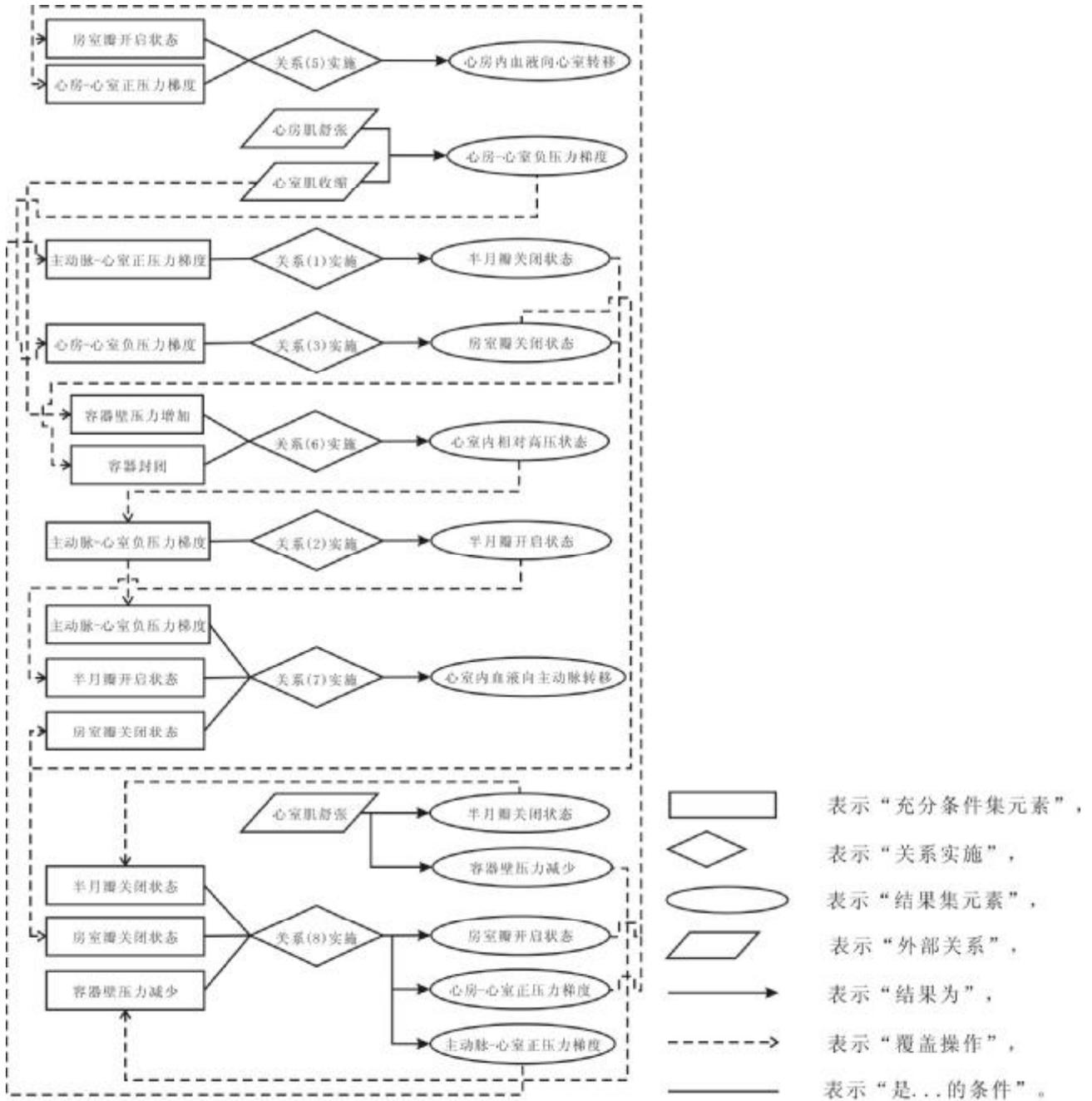


图2 心脏泵血循环中基于求供覆盖原则的一系列定性整合过程

(三) 知识的二重性：物理性和意识性

我们知道，概念是客观世界原型的在我们头脑中的反映，但这并不意味着后者原封不动地拷贝（或镜像）前者，而是经过大脑不同程度的加工处理。它们之间的关系是原象→大脑加工→模型（映象、拟象、抽象），是个由实变虚的过程。

知识的形成有二种机制，一是自然的物理机制，如物种进化，天体的运转，形成诸如动物、植物、行星、年月，一座大楼的建造等，这些事物由物理运动法则决定，反映到我们头脑中形成知识。知识的这一属性被称为知识的物理性（physical nature）。另一种是纯理性运动产生的知识，如一些数学概念和定律，包括笛卡儿坐标系、复数、一张大楼的图纸等，也包括“is_a”的抽象过程产生的概念系列，它们由思维运动的法则决定，此类知识是大脑深加工过程的产物。知识的这一属性被称为知识的意识性(mental nature)。同一个概念系列可以同时由两种机制形成（见图3，图5）。但在我们的知识储存、表达和处理时它们常常混在一起。知识的物理性运动和意识性运动是两种不同性质的运动，有着不同的规律。这两种法则在知识整合工程可能会引起“冲突”。

一篇关于UMLS语义网络的论文^[20]中指出：高层语义类型之间中的关系，一般可通过is_a连接继承给其子类型。例如，在语义类型“生物学功能”和“机体”之间存在着“是...的过程”关系，所以，“器官或组织功能”（is_a“生理功能”，后者又is_a“生物学功能”）和“动物”（is_a“机体”）之间也存在着“是...的过程”关系。然而语义类型之间的关系对属于这些语义类型的概念的所有实例之间并不一定都适合。也即这一关系在所有的具体的概念对之间可能存在，也可能不存在。例如，虽然“是...的值”关系存在于“体征”和“机体属性”之间，但一个具体的体征或一个具体的属性之间可能不存在这种关系。例如体征“超重”和“发热”可能是机体属性“体重”和“体温”的值，但“超重”不是“体温”的值，“发热”不是“体重”的值。

有些情况下语义网络中类型的位置与连接的继承会产生冲突，如果发生这种情况，那么连接的继承应被“阻断”（be blocked）。例如，按继承性，语义类型“心理功能”会是“植物”的“过程”。因为植物是无知觉的，所以这一连接显然应该被阻断。因为“关系的实质决定了其不能继承给由该关系连接的类型的子类型”。在这种情况下，“只能规定连接这两种语义类型的关系只适合这两种语义类型，不可通过 is_a 继承”。如“是...概念性部分”（关系）连接“身体系统”和“充分发育的解剖学结构”，但它不连接“身体系统”和“充分发育的解剖学结构”的所有“子类型”，如“细胞”或“组织”。

上述问题揭示了知识整合工程的复杂性。据笔者的分析，这些问题来源于上述构成知识的二种机制即物理性运动和思维性运动规律上的差异。我们知道，作为一个思维运动过程，“is_a”连接，是一种抽象过程，是概念的“必须定义属性集合（Necessary Definition Attribute Set, NDAS）”的基数减少过程，是异类实例不断加入概念的过程，是对事物的判识相对“宽容”过程。此过程认为 NDAS 的定义属性之间是相互可分的，是可以随意加减的（见图3），它并不考虑它们之间的相互依赖性、有序性或结构性。我们通过 is_a 抽象按 NDAS 演绎轴“复合体∧正常∧生命体∧脊椎∧哺乳∧随意运动 NS”→“复合体∧生命体∧脊椎∧随意运动 NS”→“复合体∧生命体∧随意运动 NS”，即把属性“哺乳”和“脊椎”逐步减去，得到机体的“哺乳动物”→“脊椎动物”→“动物”概念演绎链（见图4）。

但物理过程就不一样（见图5）。比较一下图3与图5，虽然两个图中通过两种演绎得到的概念“学生”与“医科女大学生”的 NDAS 并无区别，但用 is_a 连接的 NDAS 演绎轴和用限定关系连接的演绎轴二者反映的语义完全不同。当我们从生物进化的观点出发，按 NDAS 演绎轴“复合体∧生命体∧随意运动 NS”→“复合体∧生命体∧脊椎∧随意运动 NS”→“复合体∧正常∧生命体∧脊椎∧哺乳∧随意运动 NS”完成“动物”→“脊椎动物”→“哺乳动物”抽象时，每个 NDAS 都以前一个 NDAS 为基础，因此是个有序过程，NDAS 的定义属性不是相互可分的，不可以随意加减。因为“无脊椎动物”不可能产生“无脊椎哺乳动物”。

图6中根据 is_a 抽象过程，由 NDAS 演绎轴“复合体∧正常∧生命体∧高级 NS∧功能”→“复合体∧正常∧生命体∧功能”→“正常∧生命体∧功能”→“生命体∧功能”→“生命体”决定的概念系列“Mental

Process” → “Organism Function” → “Physiologic Function” → “Biologic Function” 中，其定义属性“高级 NS”与“心理功能”被认为是可分离的，是独立的。但“Organism” → “Plant”是一个生物进化过程，“心理功能”对“高级 NS”有依赖关系，是不可分离的，由于“Plant”的 NDAS 中不包含“高级 NS”，所以“Mental Process”不能是“Plant”的“过程”。

图 7 解释了为何“超重”不应该是“体温”的值，及“发热”不应该是“体重”的值。

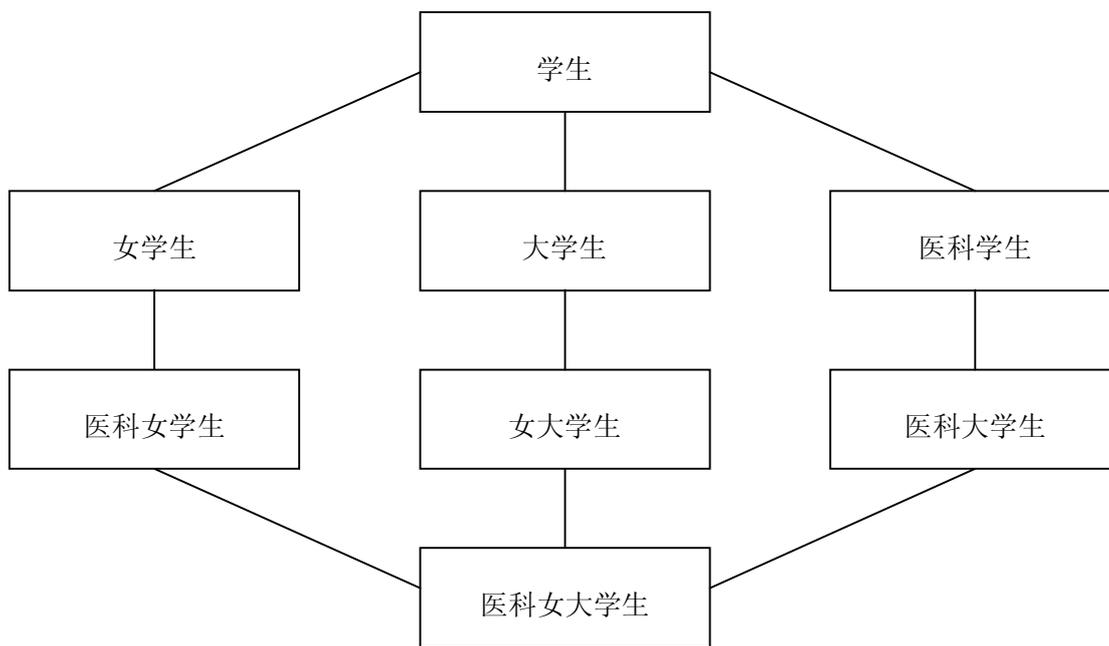


图 3 is_a 抽象过程反映了思维运动的一种规律，概念“医科女大学生”在 is_a 连接抽象过程中无需考虑 NDAS 中各属性之间的先后次序或结构，并可以沿着三条不同的 NDAS 演绎路线形成概念链。

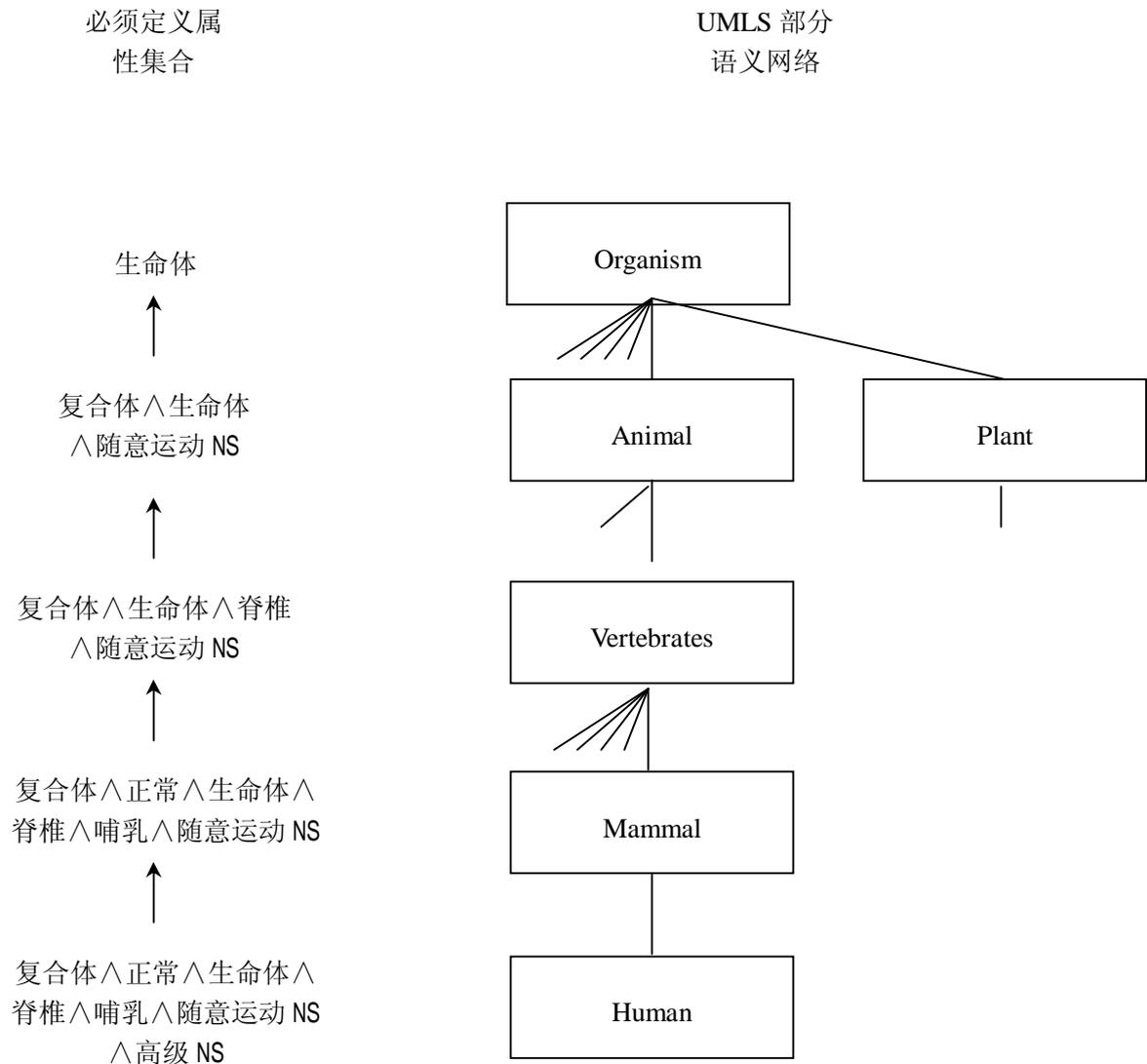


图4 “人类”、“哺乳动物”、“脊椎动物”、“动物”、“机体”的NDAS的演绎过程，而“机体”到“植物”的演绎过程并不等于其反过程，“植物”NDAS不包含“心理过程”的NDAS“高级NS”。(NS为高级神经系统)

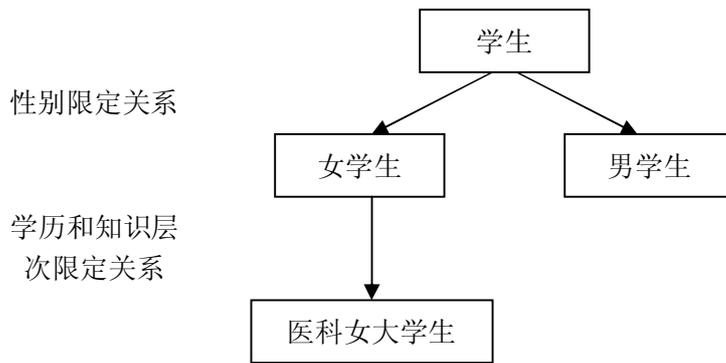


图 5 物理限定过程，反映了运动物理运动的一种规律，演绎过程也不同于图 3 的 NDAS 演绎过程的反过程。其 NDAS 大学生与专业是不可分的，所以“医科”和“大学生”需同时出现。

语义类型的必须定义属性集

UMLS 的部分语义网络

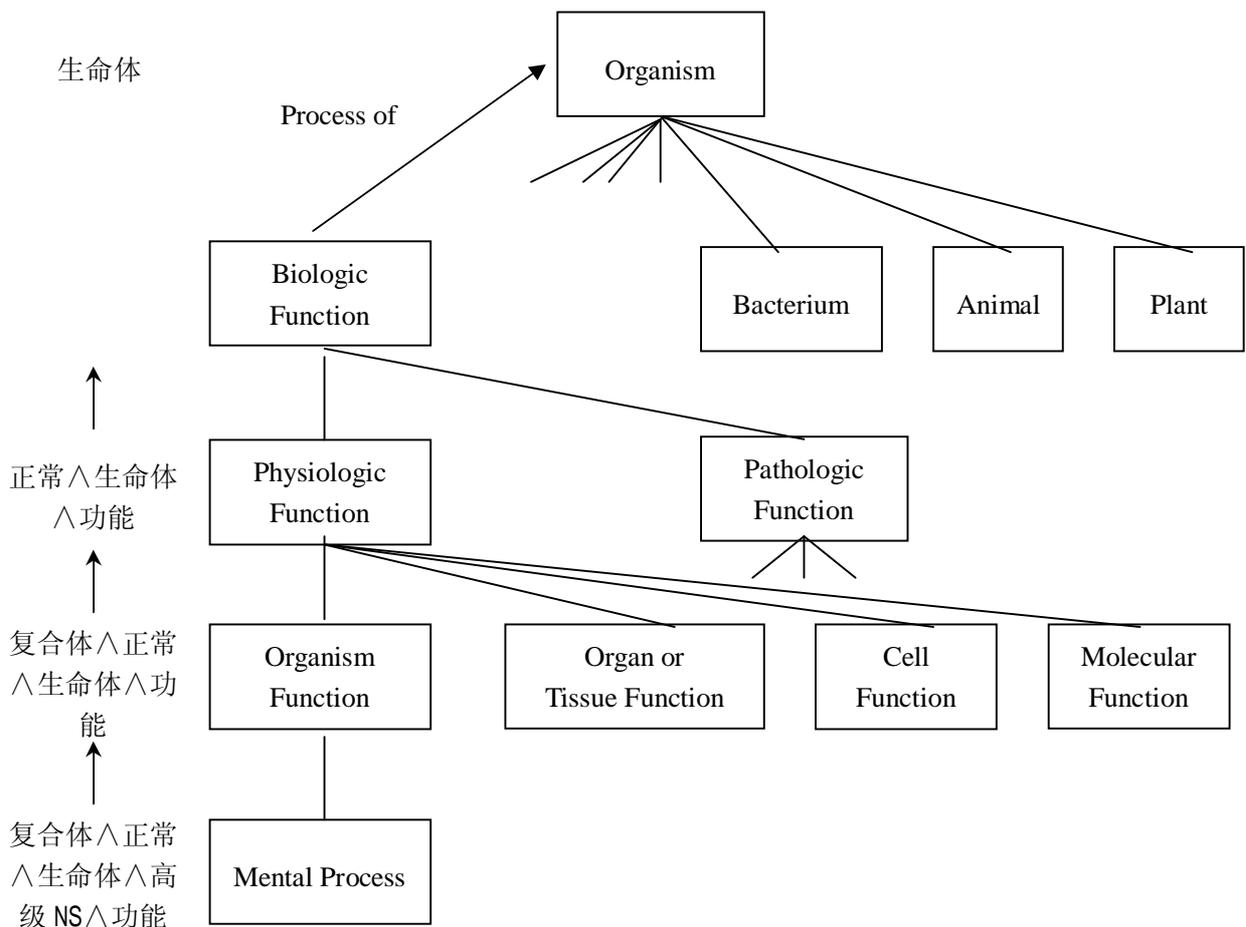


图 6 显示部分 UMLS 网络，注意随着 is_a 关系的连接链的伸展，语义类型的定义属性集的规定项目逐步减少，分类的精细程度减小。如果被减去的规规定项目与语义类型关系的有关，那么就会影响该关系的继承性。(NS 为高级神经系统)

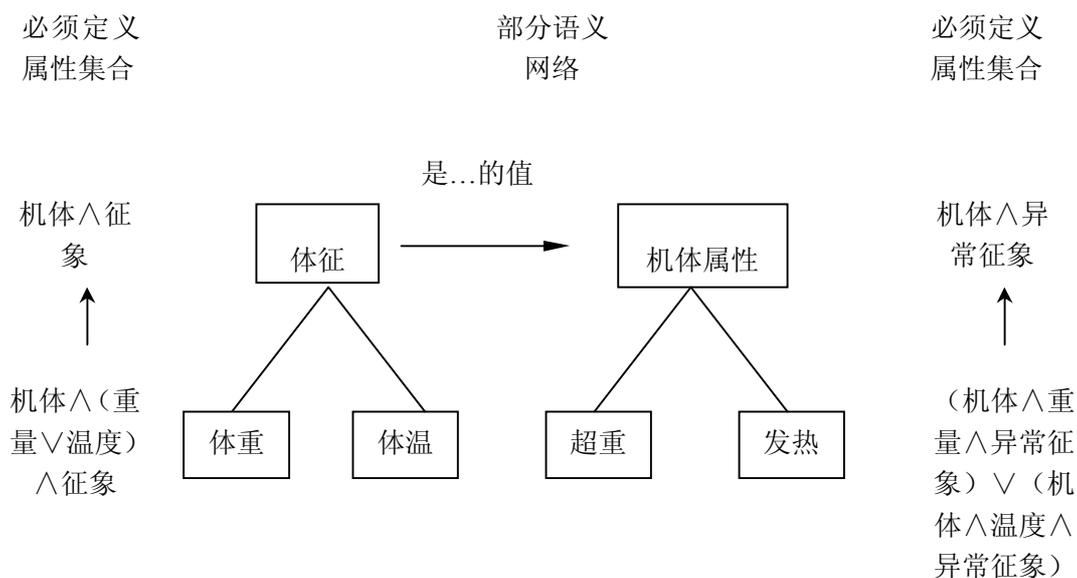


图7 由于在物理语义上，体重与超重、体温与发热分别依赖与重量和温度，但由于在思维语义 is_a 连接过程中重量和温度属性被减去，所以父概念“体征”和“机体属性”之间（上述依赖性已略去）的关系“是...值”不能为所有子概念继承。

（待续）

参考文献

1. JH van Bemmel, 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. 包含飞: 生物医学知识整合论(I), 《医学信息》杂志, 2003, 第16卷, 第6期(待发表)
3. 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
4. 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
5. Bao H.F.: The structure characteristics of the new research QMSOC and its relevant operators, J Tongji Med Univ, 1989, 9(4):235-238
6. Bao H.F.: Quantitative and Computerized Medicine--New research QMSOC(I), The proceeding of the First Conference of the Frontier for Life Sciences on Central-south China, 1989, 211-216
7. 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
8. Bao H.F.: The new functions of Quantitatively Medicine Simulating and Operating by Computer-New research QMSOC(III), 1990, J Tongji Med. Univ, 10(1): 52-56
9. 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
10. 包含飞, 刘庚妹: 计算机编码编辑技术 (HICE) 及其初级信息开发功能——新研究 QMSOC(V), 计算机应用研究 (1991 年专辑第 2 号), 1992, 4-9
 11. 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
 12. 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
 13. http://www.nlm.nih.gov/research/visible/visible_human.html
 14. <http://www.medtropolis.com/VBody.asp>
 15. <http://www.jbpub.com/visiblehuman/>
 16. http://www.education-world.com/a_lesson/lesson065.shtml
 17. 吴学谋: 《泛系——不合上帝模子的哲学》, 武汉: 武汉出版社, 1996
 18. 张镜如主编, 乔健天副主编: 《生理学》, 北京: 人民卫生出版社, 86-88, 39-45, 1996
 19. 包含飞: 续议中医学是复杂性科学——中医标准化预备研究之三 (待发表)
 20. UMLS Knowledge Sources 14 Edition, January Release, http://www.nlm.nih.gov/research/umls/UMLSDOC_2003AA.pdf