

包含飞: SNOMED的构建维度体系研究——BMKI的概念, 原理和方法学(IV), 中国数字医学, Vol(1):1, 2006, 15-19

SNOMED的构建维度体系研究 ——BMKI的概念, 原理和方法学(IV)

包含飞

上海中医药大学, 中医信息化-标准化研究室(201203)

摘要 本文以 SNOMED 的实体性术语为蓝本, 探索其深层的构成维度体系, 试图为未来新一代的医学智能系统的“认知意向”能力的建立寻找依据。

关键词 医学信息学 生物医学知识整合论 知识工程 智能系统 认知意向

The Research on The Constructing Dimensions of SNOMED ——The Principles, concepts and methodology of BMKI(IV)

Hanfei Bao

The Lab of Info and Stan of TCM

Shanghai University of Traditional Chinese Medicine(TCM) (201203)

Abstract This research takes the sections of entity terms of SNOMED as the object to analyze its fundamental dimensional system, in order to explore how to equip the new generation of medical intelligent systems with the function of “cognitive intent”.

Key words Medical Informatics the Theory of Biomedical Knowledge Integration(BMKI) Knowledge Engineering Intelligent System Cognitive Intent

一, 生物医学知识工程的理论幼稚病

随着诸如知识领域本体(Domain Ontology), 本体 web(Ontology web)等人类知识工程的兴起, 生物医学信息学的知识工程又一次向令人望而生畏生命系统发起冲击, 又一次向浩如烟海的人类认知本身发起冲击。人们开发了术语级知识标准如 SNOMED 和事件级(或称消息级)的标准如 HL7, 不断地完善和增强它们的功能; 人们还正在开发各个领域的标准的概念层次的模型即本体; 人类更雄心勃勃地开展全世界范围的合作, 开发本体 web(或语义 web); 生物医学知识整合论(BMKI)则在较彻底研究了生物医学知识全局以后, 指出标准的概念层次的知识本体(即概念性本体, Conceptual Ontology, CO)不足以实现生物医学知识运算层次的大整合, 必须同时建立实验性知识本(Experimental Ontology, EO)。^[1-3]

与以前或多或少带有盲目自信，或多或少有点草率从事的做法有所不同，这次冲击较为严谨和审慎，对基本问题思考较深，战略上带有稳扎稳打、边试边干的色彩。上世纪 80 年代末到 90 年代初那样的野心勃勃地热衷于开发“具有大脑思维功能”的“第五代计算机”的过热景象已经不复存在。

但这并不是说当今的生物医学知识工程就不存在危机。相反，当前的知识工程的成功主要在较低级的信息语法层次（数据格式，数据组织，数据协议，数据传输，数据显示等）。至于再高的层次的知识工程，别说是知识的语用层次，就是知识的语义层次，就是对人类智慧的严重挑战。只要涉及知识的语义层次操作，知识工程就必须十分谨慎。

某种程度上讲，当前生物医学知识工程仍带有盲目性弊病。笔者并未见到平行的甚至先行的关于知识工程的较为深刻的理论探索及其实质性的进展，好象只要一味努力地进行规范化和不断地进行资源积累，我们美丽的梦想一定会变成现实似的。

笔者认为如果缺乏对生命系统和人类自身的认知和思维功能的深层次的本质性的了解，那么对当前的生物医学知识工程前途仍不能高枕无忧。当我们的智能系统试图对超级复杂的生命大千世界进行运算、推理时，我们能“胸有成竹”吗？如果是这样，那么“成竹”在何方？如果不是这样，那么我们“胸有成x”？

二，生物医学智能系统的认知意向何在？

一个真正的智能系统在面对它要处理的对象时，除了必须拥有必要的知识背景以外，还必须具有某种“认知意向”。打个通俗的比方，一个人挑选对象，你是着重他或她的外貌？，还是才能，品德？这就是认知意向。不同的“认知意向”决定认知主体从不同的视角来审视对象及他的取舍标准。这种观测事物的视角就是认知科学上的维度（或基维度）。

相对于物理世界的物质运动的参量如方向、轨道和速度等，认知维度是人类大脑运动（认知、思考和探索）的方向、路径或轨道，它们决定了意识或心智运动的方向。可以认为“认知意向”由基维度集合或其框架组成。“认知意向”框架的确立对该知识领域或知识体系的智能系统的数据挖掘或决策支持的操作、处理、运算的始法、取向、择道等具有全局性指导意义。

那么，如何给生物医学智能系统赋以具有“本能”的“认知意向”？如何使医学智能系统在不同的认知任务下形成不同的“认知意向”（趋向，倾向，惯性）？如何使我们临床医学智能系统拥有医生的“职业思维”习惯？

这是生物医学知识工程所面临的不可回避的重大课题。不解决这个问题，就不能充分发挥智能系统超常运算速度和超大的储存容量的优势。

一个切实可行的方法是选择一个较典型和严谨的医学知识体系，深究隐藏在其背后的医学专家的认知和思维基本取向和轨迹，分析其认知基维度的（静态的、动态的）框架，指导建立我们的医学智能系统“认知意向”的元结构。这就是说，对一个知识体系或知识领域赖以成型的维度体系进行全面的、系统的、结构性分析，有助于发掘和构建其“认知意向”框架。

BMKI 拟对医学智能系统的“认知意向”课题作深入探讨。本文首先从当前最为成功的医学术语学标准知识 SNOMED 开始。

作为一个知识体系的背景理论的探索，我们无须追踪其最新或最完整的版本，应用 Roger A Cote 原著，李恩生主译的英汉对照《国际医学规范术语》（精选本）（1997 年版）（以

下简称 SNOMED）足以满足我们的学术探索需求。

三，构建 SNOMED 的认知维度分析

复合为万物万机之源，生物医学系统是经历漫长进化过程发展而来的高度复杂的系统。对生物医学体系的认知和探索，无疑是对人类智慧的极大的挑战。为了适应客观世界的高度复合性，知识世界或意识世界同样通过概念复合而导致千变万化的概念体系。SNOMED 的体系结构反映了这一规律。生物系统及其知识体系的多样性或复杂性，主要受以下原始物理机制和意识机制驱动：

- (1) 客观物理世界的复合导致多样性（**物理多样性机制，机制1**）；
- (2) 对应的认知世界的复合导致多样性（**意识多样性机制，机制2**）；
- (3) 认知世界自身对客观世界的重组织（**意识重组多样性机制，机制3**）；
- (4) 认知目标意义下世界的多样性（或异同性）（**认知目标多样性机制，机制4**）；
- (5) 功利目标意义下世界的多样性（或异同性）（**功利目标多样性机制，机制5**）。

SNOMED 包括**实体**（概念实体和物理实体）和**关系**二部分。

实体部分包括（1）机体的正常物理可见实体（局部解剖学 T）和（2）形态学 M）；（3）（肉眼）不可见实体（功能 F）；（4）异常物理可见或不可见实体及其认识（疾病/诊断 D）；（5）人类对物理实体的控制和干涉过程（操作 P）及（6）物质（化学制品、药物和生物制品 C），（7）人体以外的其他机体(L,如病毒，细菌，动物，植物等)；（8）可对上述因素发生作用的物理因素、活动和力(A)以及（9）其他环境职业(J)和（10）社会环境（S）。

关系部分为把上述概念关联或复合起来成为更复杂的概念的关联词/修饰词（G）。

全部内容按字母和数字（16 进制）组成的编码排列起来。

下面先就 SNOMED 的实体部分的重要内容对其进行维度体系分析。

3. 1 局部解剖学部分

SNOMED 的局部解剖学（Topography）主要由物理性空间结构关系和抽象-具体关系组成，其中：（1）物理性空间结构关系由**机制1**驱动：如比邻（物理界定，人为划分）关系（**—*is adjacent to*—**），局整关系（**—*a part of*—**）等；（2）抽象-具体关系（**—*is a*—**）由**意识机制3**驱动。其意识运动性质，认知层次，推导功能见表 1。

表 1 抽象关系与物理结构关系的认知科学分析。

语义关系性质	认知驱动	认知层次	推理功能
抽象关系 (<i>—is a—</i>)	心智求同 运动	理解	属性继承
物理结构关系 (<i>—a part of—</i>) (<i>—is adjacent to—</i>)	感受与观测	描述	空间转运运算等

局部解剖学的概念绝大多数由比邻（物理界定，人为划分）关系等形成。由抽象关系（*—is a—*）生成的概念有“身体组织”，“供者部位”，“受者部位”，“腹腔内脏”等，为意识运动对事物的共同属性的提取即内涵的认识，并产生对物理世界的分类或组织。

3. 2 功利目标多样性机制驱动的概念分类

医学作为防病治病的实用科学，最重要的认知目标维度是对象的**危害性程度和可控程度**，医学上的良恶性实际上这二者的**复合维度**（凶险而可控，或难控而非凶险都不是恶性的，只有凶险而难控才是最恶性的）。由分类意义的最高的判据即**机制5** 驱动。例如肿瘤的第 5 位为肿瘤与医学控制目标有关的肿瘤行为的编码：良性（标号为 0），良恶未定或良恶交界（标号为 1），原位癌（标号为 2），恶性，原发部位（标号为 3），恶性，继发部位（标号为 6），恶性，原发继发部位未定（标号为 9）。肿瘤的第 6 位为肿瘤与医学控制目标有关的良恶性级别有关的编码：良性（标号为 1，I 级），中度分化（标号为 2，II 级），不良分化（标号为 3，III 级），未分化（标号为 4，IV 级），难以分级（标号为 9）。

3. 3 各种机制的复合

绝大部分医学术语是**机制1, 2, 3, 4, 5**的合理复合而成,但所有的复合必须尽可能满足**机制5**(功利机制)。下文描述 SNOMED 术语的构成维度,对大部分维度及其复合尽可能列出 SNOMED 的术语举例(用斜黑体表示), $\wedge$ 和 $\vee$ 分别表示“与”和“或”复合运算。

3.3.1 一般系统的认知维度:结构,功能和属性:

SNOMED 的这类术语能明晰地抽象出它们深处的**认知维度**。这类维度的举例有空间位移历史:**原发-继发**;频发-贫发:**典型-非典型**;局整性:**完全-非完全**;表里性:真性-假性:**假性肥大**;空间分布性:**弥漫性,局限性**;过程速率:**急性-慢性**;体积:**肥大,巨型肥大**;中央性-周边性:**向心性肥大**;多侧性:**单侧肥大**;虚实性(如**体重丧失**)。

3.3.2 可分析的认知维度及其复合所致概念体系的多样性:

3.3.2.1 物种进化所致概念体系的多样性:如**退化**。个体繁殖(遗传):如**先天畸形**。

3.3.2.2 生物功能和属性所致概念体系的多样性:生长速率:**生长迟缓,生长停止,生长改变,增生**等;应激性,如**应激增高**;中毒性;带菌性;共生性;应感性:如**易感性**;耐受性:如**耐受不良**;壮健性:如**衰竭,危象**;(神经或体液)调控性:如**失禁**;适应性;恒温性:如**不明原因的发热**;强弱度:如**功能不全**;生长与发育 $\wedge$ 更新代谢性:**骨更新率障碍**;生物性 $\wedge$ 频发-贫发:**非典型性增生**等。

3.3.2.3 生物结构所致概念体系的多样性:细胞结构:**细胞和亚细胞层次**。

3.3.2.4 生物机体整体的功能所致概念体系的多样性:如**心理或性功能**。

3.3.2.5 多维度复合所致概念体系的多样性:

这类维度复合如生物过程 $\wedge$ 空间位移:**原发性增生**;物理对象 $\wedge$ 病理过程:**骨肥大,骨刺**;形态 $\wedge$ 病理过程:**结节性增生,腺瘤结节,息肉状结节,乳头状增生,囊性增生,棘皮瘤**等;性质 $\wedge$ 病理过程:**间皮细胞性增生,基底细胞性增生**;性质 $\wedge$ 形态 $\wedge$ 病理过程:**灶性上皮细胞性增生**。

SNOMED 此类术语从观测属性又进一步有如下分类:

宏观可辨 $\wedge$ （结构 $\vee$ 系统 $\vee$ 功能） $\wedge$ （正常 $\vee$ 异常）类。肌肉骨骼系统：如**骨形成，骺形成障碍**；呼吸系统：如**咳嗽，咯血，窒息**；心血管系统：如**正常血压，心功能，心功能减弱**；体表系统：如**正常角质化，角质化增加**；消化系统：如**胃酸度异常，呕吐**；泌尿系统：如**异常肾功能，遗尿**；生殖和围产期：如**男性不育，异常月经，痛经**；神经系统；特殊感官等。

微观可测 $\wedge$ （组织 $\vee$ 结构 $\vee$ 系统 $\vee$ 功能） $\wedge$ （正常 $\vee$ 异常）类。造血系统：如**血小板产生增加，溶血**；免疫系统：如**免疫学活性，免疫抑制**；激素-维生素-受体部位：如**异常激素生成，旁分泌物质**；遗传-分子-细胞。

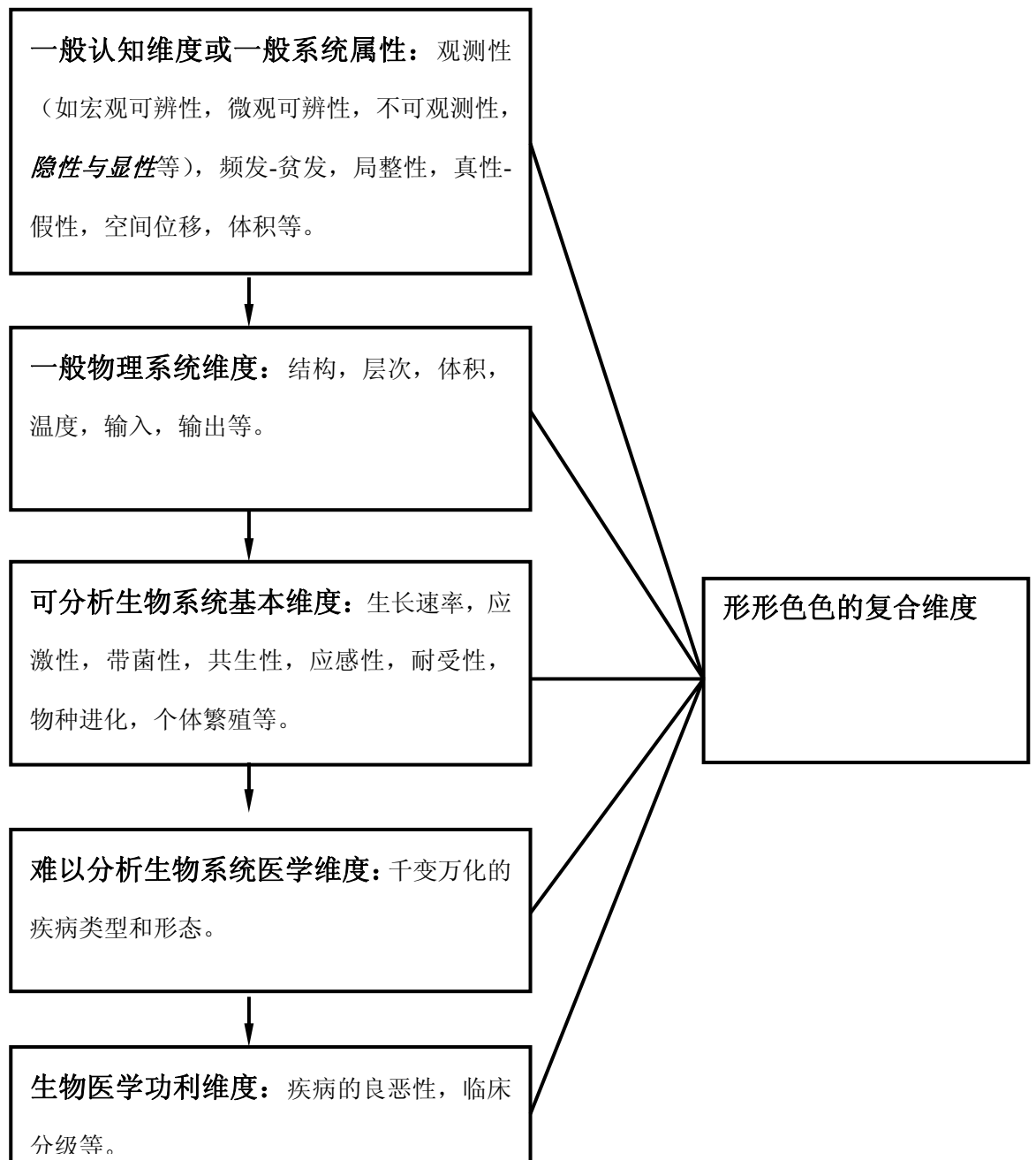


图 1 认知性维度及各层次的物理维度的关系。

疾病/诊断的复合维度的形式为： $(\text{宏观可辨} \vee \text{微观可测}) \wedge ((\text{物质} \vee \text{亚细胞} \vee \text{细胞} \vee \text{组织} \vee \text{结构} \vee \text{系统} \vee \text{整体}) \vee \text{功能}) \wedge \text{异常}$ 。根据病因等因素又增加了先天性疾病，代谢和

营养紊乱，精神障碍，损伤与中毒，传染和寄生虫病。

3. 3. 3 难以分析认知维度所致概念体系的多样性：

由于生物机体系统的高度复合性，任何复合失调均可能导致新的病变品种，导致生物机体的异常或病变同样呈现极度的多样性。例如几乎任何组织必须与血管系统复合，因而几乎任何组织都有可能产生出凝血疾病（如肺部咯血）。但人类当前的认知能力尚不足以全面清晰地分析和区别如此众多维度的复合，我们不得不根据生物哲学层次理解它们并统归于生物复合所致的多样性一类。例如人类机体的疾病类型较其他任何系统更丰富多样，SNOMED 的大部分实体性术语属于此类：**感染，脓皮病，湿疹，硬皮病，皮饥炎，纤维化，息肉，溃疡**等。炎症类型也可细分为**血管样炎症，苔癣样炎症，脂溢性皮炎**。

对此类 SNOMED 术语我们只是通过“具体的生物学观测方法”发现和描述所有这些类型，只能无条件地认可这些由**认知现象**决定的类型或多样化，却难以象上节（3. 3. 1 节）一样明晰地抽象出它们深处的**认知维度**。

疾病/诊断方面术语是 SNOMED 的核心部分，受功利机制（**机制5**）驱动，术语往往须指定物理主体（物质、亚细胞、细胞、组织、结构、系统、整体）。我们这些物理主体是由历史悠久的生物进化（被称为自组织过程）形成，结构纷繁复杂。这种高度复合性也造就了大量的疾病主体术语，对此类术语我们也只能从**认知现象**而难以从**认知维度**区分其类型，认识其多样性。

3. 3. 4 SNOMED 维度体系的层次和复合

分析上述各种维度及其复合规律，可给出各种维度类型及其关系图（见图 1）。箭头由较基础的维度指向更高级的、更复杂的维度。

四，结束语

生物医学智能系统的“认知意向”的建立依赖于人类在该领域的认知维度及其体系结构，本研究从 SNOMED 实体性术语体系构成规律出发，初步勾画出其构成维度体系。进一步的

工作是分析 SNOMED 的关系（连接词/修饰词）部分的术语体系构成规律，以期得到更为完整的构成维度体系。为建立生物医学智能系统“认知意向”提供依据。

参考文献

1. 包含飞：概念性本体、实验性本体和罗盘-灯塔式本体——BMKI的概念，原理和方法学（I），《计算机应用与软件》，22（增刊）：113-115，2005
2. 包含飞：SNOMED的一般连接词/修饰词的元维度及其复合研究——BMKI的概念、原理和方法学（II），中国医院，22（增刊），113-115，2005
3. 包含飞：物数学及其形态数——BMKI 的概念，原理和方法学（III），数理医药学杂志（已接受，待发表）
4. 包含飞. 生物医学知识整合论（I-X），医学信息学，2003-2005
5. Hanfei Bao: The Theory of Biomedical Knowledge Integration(BMKI)（I-VII), Medical Information, 2003-2006
6. Roger A Cote 原著，李恩生主译：英汉对照《国际医学规范术语》（精选本），北京：北京医科大学，中国协和医科大学联合出版社，1997
7. 陈世福，陈兆乾等编，《人工智能与知识工程》，南京：南京大学出版社，1997

