

包含飞：《计算机应用与软件》，22（增刊）：113-115，2005

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）

包含飞

上海中医药大学中医学信息化-标准化研究室

201203, bhf2002@online.sh.cn

摘要 本文从分析信息单元的三个组成部分信息载体、变量和变量值开始，演绎了群体知识与个体知识之间的转化规律。分析了各种类型的变量及其特性。本文把当前作为人工智能研究开发热点的本体称为概念性本体，根据其标准性、群体性、共性性而引导出另一个具有本体性质（标准性、群体性、共性性）的涵盖广大的知识本体，即实验性本体，并把二者结合而引申出罗盘-灯塔式本体。最后又进一步扩展到个体性知识而给出 BMKI 的基本公式。

关键词 人工智能 医学信息学 生物医学知识整合论 医学本体学

本系列将对生物医学知识整合论(The Theory of BioMedical Knowledge Integration, BMKI)的原理、概念和方法学作更为详尽的、具体的探讨。

1. 对信息体的分析

从根本上讲，任何信息单元需要回答三个问题：（1）我们观察的对象是什么？（2）因为一般来说一个对象有无穷多的方面，那么我们关注的是对象的哪一方面？（3）观察的结果如何？我们可把这三部分分别称为广义的信息载体 C（观测对象，Carrier），变量 F（观测方面,Feature）和变量值 V（观测值，Value），三者组成一个信息单元 U，因而 $U = (C \times F \times V)$ （见图 1）。变量 F 被 SNOMED 称为 observable,解释为“Entities that are the subject of an observation（作为一个观察行为的承受者的实体）”。

信 息 载 体 (Carrier)	变 量 (Feature)	变 量 值 (Value)
----------------------	------------------	------------------

图 1 信息最基本单元（U）模型,包括载体（Carrier），视点或变量 (Feature)和变量值或观测值（Value）。

信息载体可分为个体与群体，由此知识可分出个体知识（信息）和群体知识（信息）。二者的定义如下：

(1) 定义：

定义 1: 个体知识 关于某一实例的知识（和数据），包括对其静态或动态特征及与其他事物的关系的描述。

定义 2: 群体知识或标准知识 关于某一类事物共性的知识（和数据），包括对其静态或动态特征及与其他事物的关系的描述。

(2) 个体知识与群体知识的转化:

在人类的认知过程中，我们常常进行上述二种知识的转化。

a. 我们知道，思维抽象化过程是对某些个体属性的舍弃而对某些共同属性的保留过程。以 $\text{—}a\text{—}$ 和 $\text{—}s\text{—}$ 分别表示**抽象化**为关系和**具体化**为关系：

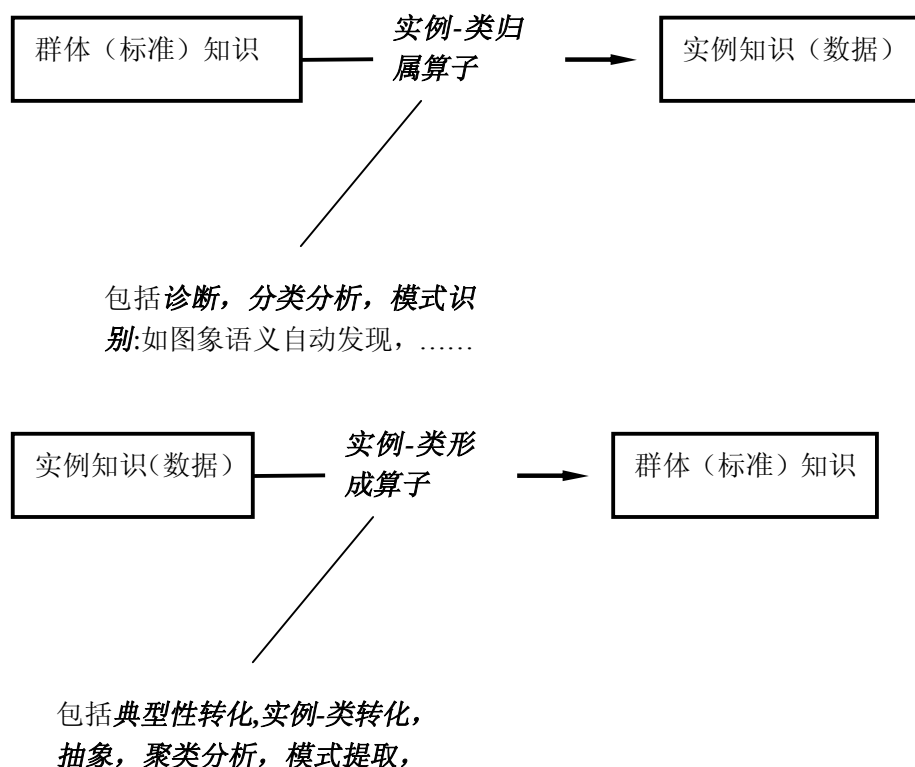
如 A 类和 B 类的关系为 $A\text{—}a\text{—}B$ 或 $B\text{—}s\text{—}A$ ，则称 A 为 B 的子类或实例，B 为 A 的父类或上级类，关于 A 的知识相对于关于 B 的知识为个体知识或具体知识，而关于 B 的知识相对于关于 A 的知识为群体知识或抽象知识。

同样地，命题的一般性和具体性也存在这种关系，图 2 为相对一般性命题与相对具体性命题之间此种转化关系的示意图。



图 2 一般性命题与具体命题之间的转化。

b. 知识模式的发现和指定: 模式发现是个体知识向群体知识的转化，求群体中相对不变的共性的过程；模式指定是模式发现的反过程。是在群体知识对个体知识的指导下，考察个体特性提供的限定或约束集合，从而得出个体归属于（或诊断为）何种模式或类型的过程。两种方向的互相转化见图 3。



(3) 信息单元成分之间的其他运算:

我们可以依据信息单元及其组分之间的-is a-关系, -a part of-关系, 代数关系, 函数关系, 映射关系及各种物理语义之间的关系对信息单元进行操作或运算, 也可通过加上必要的约束使事物由相对一般、相对粗粒度、相对不确定向相对具体、相对细粒度、相对确定的方向转化。

3. 变量类型

上文说过, 所谓信息单元的变量就是我们的观察事物的内容或方面。事实上变量是个色彩斑斓的世界, 选择何种变量是认知过程实施者的智慧的表现。但即便如此, 我们仍然可以对其作一大致的分类。

说到底, 任何信息单元的任何变量都是一种关系, 即一种二元或多元关系, 包括多元序关系。我们按照信息变量的惯常性与特殊性、恒定性与变化性、单纯性和复合性可把变量分为静态变量和动态变量。前者包括特性(或状态)变量及关系(或势态, 明的潜的)变量, 后者可看作静态变量的连续有序复合系列。

(1) **特性变量**往往指一些相对固定不变的属性, 又可称为**常量类型**, 其变量值往往可以用常数值或一个恒态描述表示, 包括数值型, 如测眼压(修兹眼压计)右眼 6.6kPa (49.76mmHg), 左眼 2.7kPa (22.55mmHg), 及文字型描述表示, 如(球结膜)充血, (角膜)混浊。其形式为 $\&f$, 其中 $\&$ 为赋值算子。

特性变量中有些关系的参与方是不言自明的或缺省的, 如我们无需把“高度”描述为“对象与尺子的关系”, 把“颜色”描述为“对象与照射光的关系”等等。

(2) **关系变量**较多是二元关系, 如左右眼瞳孔是否**等大**, 上穹隆是否**存留**异物, 外耳道**注入**冰水(50ml)后的反应, 碱性药物**治疗**腹痛, 瞳孔的对光**反射**, “视觉中对红色的**分辨能力**”等。其形式为 $\&(a_1*a_2)$, 其中 a_1, a_2 为参与关系的方面, $*$ 为关系, $\&$ 为赋值算子。

但实际上任何事物的关系变量是“潜无穷”的, 视观察者施加的另一激发因素(或方面)而定。因此我们又称它为一种“势态”或“潜质”。例如一个司机的“**反应能力**”往往要看他对某种“意外交通状态”如一个孩子, 一条狗, 或一辆自行车的突然出现或刹车突然失灵情况下的应变能力才能判断, 平时难以观察这种潜能。

(3) **动态关系或过程**是一个关系链被启动并得到实现的过程。其形式为 $\&(r_1*r_2*...*r_n)$, 其中 $r_i=(a_1*a_2*...*a_n)$, $a_1, a_2, \dots, a_n$ 为参与关系的各个方面, $\&$ 为赋值算子(可以是对过程的描述)。

动态关系型变量的值一般不是一个常量, 可以用函数表示的一种势, 如 $x = 2(x)$ (见图4)。

1	2	3	4	5	6	7...
2	4	6	8	10	12	14...

图4 函数关系表示一种势或场, 二个关系方面的相对运动变化的趋势。

4. 概念性本体、实验性本体和罗盘-灯塔式本体

T. R. Gruber 指出^[11] 指出,“本体(ontology)一词在人工智能中引起很多争论,哲学中这一单词已有很久的历史,是‘关于存在’的研究(the subject of existence)。在认知科学中本体是指“关于知识和认知过程”(about knowledge and knowing),也常常导致困惑。在知识共享研究领域中,我用术语本体表示一个概念化过程的规定。所以本体为应用于代理(agent)或代理群体对一组概念和关系的描述(象对一个程序的规划的正式规定一样)。这一定义与把本体当作‘一组概念定义’是一致的,不过更为一般化而已。因此它确实与该术语在哲学中的意义不同。”

“重要的是清楚建立本体的目的。我与我的同事开发本体的目的是知识共享和重复使用。在这一宗旨下,本体是用于实施本体性受托任务的一个规定。本体性受托任务的正式定义由下文给出。从实际出发,我们选择把本体写成知识词汇的一组定义。虽然这并不是规定一个模型化项目的唯一方法,但已经在 AI 软件的知识共享中具有良佳的特性(例如对读者和语境有语义独立性)。事实上,一个本体受托任务是应用某个词汇的一个协议(即提出疑问和作出断言),要求在由本体规定的理论方面以一致的(但不一定是完全的)方法实现任务。我们建立代理,它们把任务提交给本体。我们设计本体使可以在代理之间共享知识。”

按笔者的理解,既然本体的重要特性是共享性和重复应用性,那么**本体首先必须是关于标准的(典型的、共性的、群体的)概念性知识**。基于对本体的这一本质的理解,笔者认识到另一方面的标准的(典型的、共性的、群体的)知识群体,即实验性的或映射性或函数性知识群体也同样具有“本体性”,并把它们称为实验性(或非典型性)本体(Experimental Ontology, EO)。并把作为当前的知识库开发热点的 AI 界共识的本体相应地称为概念性(或典型性)本体(Conceptual Ontology, CO)。

定义 3: 概念性知识本体为意识性或理性模型化项目,即某领域的概念结构模型。

定义 4: 实验性知识本体为物理性或映射性模型化,包括实验性数字映射或函数关系等。

概念性知识本体是当前人工智能、认知科学、医学信息学的知识工程的研究和开发的一块热土。借助一些新型的本体语言如 DAML+OIL、OWL(Web Ontology Language)等开展了分布式、结构化(主要是框架式)、可共享的、可重复应用的开发,有全球大联合的发展趋势^[12-13]。我国有些单位也在思考参与一些具有互补意义的内容的研究和开发^[14]。

与此形成强烈反差的是,实验性知识本体仍是一片巨大的未开发利用的处女地。属于笔者曾提议开发的诸如“体内世界,星转斗移”标准数据人体^[7]的内容。如果把概念性本体看作是“知识工程大厦”的“钢筋混凝土构架”或一位“骨架美女”的话,那么实验性知识本体就是不可或缺的补充,否则“大厦”难以完美而突显其雄姿,人们也难以见到一位“长袖善舞”的美人。本文拟把 EO 纳入整个生物医学知识工程之中,并把 CO 和 EO 二者整合为 BMKI 的 CBO 总战略,即**罗盘-灯塔式本体**(Compass-beacon Ontology, CBO)^[6],以克服当前生物医学知识工程“天阙一方”的遗憾。

定义 5: 罗盘-灯塔式本体为 CO 与 EO 的结合。

我们已指出,本体性知识既然是标准性知识,那么它们必然是关于群体的知识(knowledge of population, KP)。而电子病历、关系性数据库等知识大量是个体性知识(knowledge of individual, KI),如此整个生物医学工程各基本领域之间的关系如图 5 所示。

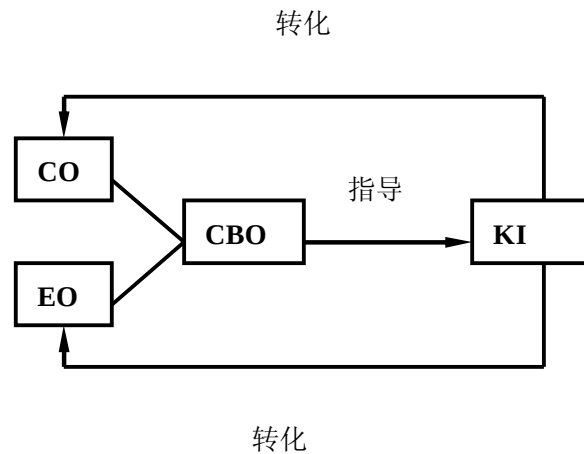


图 5 概念性本体、实验性本体及其复合罗盘-灯塔式本体及个体知识的关系示意图。

4. BMKI 基本形式

由上文所述，BMKI 的任务是探索群体性知识概念性本体、实验性本体及个体性知识三大部分（见图 6）的关系、相互转化和连接。其基本形式则为 $BMKI = \Theta[(CO \oplus EO) \otimes KI]$ 。此处 BMKI 为生物医学知识整合论（学），CO 为概念性本体，EO 为实验性本体，CO+EO 为群体性知识（KP），也即 CBO，KI 为个体性知识，符号 $\Theta, \otimes, \oplus$ 为关于关系、转化和连接的整合算子。

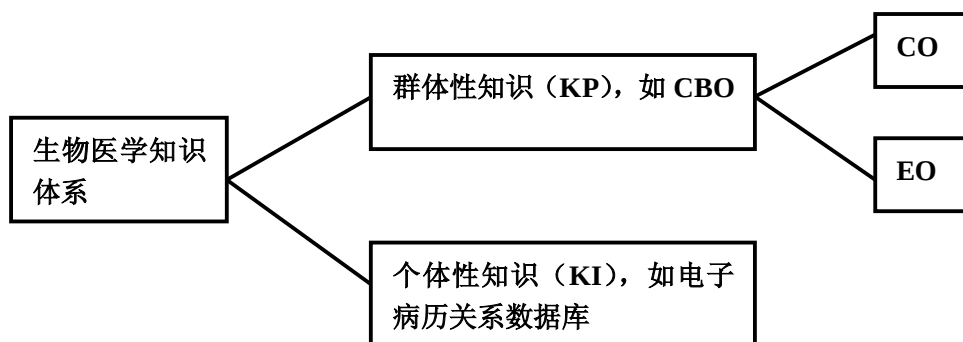


图 6 生物医学知识体系的三大部分。

- 1.包含飞: 生物医学知识整合论(I),《医学信息》杂志, 16(6): 174-279, 2003
- 2.包含飞: 生物医学知识整合论(II),《医学信息》杂志, 16(8): 410-415, 2003
- 3.包含飞: 生物医学知识整合论(III),《医学信息》杂志, 16(11): 602-605,2003
- 4.包含飞: 生物医学知识整合论(IV),《医学信息》杂志, 16(12): 666-670,2003
- 5.包含飞: 生物医学知识整合论(V),《医学信息》杂志, 17(5): 244-250,2004
- 6.包含飞: 生物医学知识整合论(VI),《医学信息》杂志, 17(8): 452-457,2004
- 7.包含飞: 生物医学知识整合论(VII),《医学信息》杂志, 17(11): 685-692,2004
- 8.包含飞: 生物医学知识整合论(VIII),《医学信息》杂志, 18(1): 11-15, 2005
- 9.包含飞: 生物医学知识整合论(IX),《医学信息》杂志, 18(3): 172-176, 2005
- 10.包含飞: 生物医学知识整合论(X),《医学信息》杂志, (待发表)
11. T. R. Gruber.: What is an Ontology?
<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>
- 12.梅婧 刘升平 林作: Chpt-63-语义Web 语言的逻辑分析,
<http://www.is.pku.edu.cn/~lz/publications/semanticweblang.pdf>
- 13.M. Taboada1, et al, 包含飞, 王晓幸翻译:DAML+OIL 应用实例: 眼科学领域的一个本体,<http://www.miforum.net/bhf/ykbt.doc>
- 14.王晓幸等: 眼科信息学: 数据与知识的整合研究(I), (待发表)