

# Part-Whole 关系的细分及性质分析

连莉<sup>1</sup>, 马军<sup>1</sup>, 雷景生<sup>2</sup>, 宋玲<sup>1,3</sup>

(1. 山东大学计算机学院, 济南 250000; 2. 海南大学信息科学技术学院, 海口 570228; 3. 山东建筑大学计算机系, 济南 250000)

**摘要:** Part-Whole 关系的语义差别往往决定了本体是否可以合并或重用。Wiston 分类法被用来对医学本体中的 Part-Whole 关系进行了分类, 使每类 Part-Whole 子关系具有特定的语义并有传递性。利用 Mereology 理论分析了子关系的性质和子关系之间的相似性, 以及子关系之间相似性和逻辑推理正确性的关系。为 Part-Whole 关系的研究提供了新的方法学。

**关键词:** 本体; Part-Whole 关系; 语义 Web

## Subdivision and Nature Analysis of Part-Whole Relation

LIAN Li<sup>1</sup>, MA Jun<sup>1</sup>, LEI Jingsheng<sup>2</sup>, SONG Ling<sup>1,3</sup>

(1. Department of Computer, Shandong University, Jinan 250000; 2. School of Information Science and Technology, Hainan University, Haikou 570228; 3. Department of Computer, Shandong Architecture University, Jinan 250000)

**【Abstract】** The difference between the semantic meanings of Part-Whole(PW)relations often determines if the ontologies can be merged or reused. Wiston taxonomy is employed to partition the PW relations in medical ontologies into sub-relations, such that, each sub-relation has special meaning and satisfies the transitivity. Mereology theory is used to analyze the properties of the sub-relations, the similarity between the sub-relations and the correctness of the reasoning based on PW relations. The research results provide new methodology for the study of PW relations.

**【Key words】** Ontology; Part-Whole relation; Semantic Web

Part-Whole 关系是最常用的包含关系之一。目前,人们常常不考虑 Part-Whole 关系的具体语义,而只是把它作为一种广义的部分-整体关系来表示层次结构的领域知识。但即使对生物学、医学等同一应用领域的本体,Part-Whole 关系所表示的部分-整体关系语义可能有着本质的不同。若不进行细分,不仅不能正确地理解由本体等表示的领域知识,而且在本体的重用和合并后,在新的知识表示中,很可能造成 Part-Whole 关系所表示的语义不一致。

另外,一般 Part-Whole 关系不满足传递性,而传递性往往是构成逻辑推理的基础,所以对 Part-Whole 关系细分的另一个原则是使每个子关系具有传递性,使得可用传统的谓词逻辑进行推理和基于语义检索等。

在相关的研究工作中,Winston和Odell<sup>[1]</sup>把Part-Whole关系分成6类。文献[2~6]的研究主要涉及对具体领域中对Part-Whole关系的分类和应用。研究的重点在本体的建模和实现,并没有用形式化方法对不同类型的Part-Whole关系进行性质上的刻画和理论分析。

本文讨论了对Part-Whole关系的分类和子关系之间的性质。并进一步分析了子关系的划分以及和推理正确性之间的统计关系。为了避免过于抽象的形式化讨论,本文选用在医学本体构造中所使用的Part-Whole关系。首先利用Wiston分类法对医学本体建设中所使用的Part-Whole关系进行了分类,并进一步采用Mereology理论系统,分析了子关系之间的性质和子关系之间的相似性,讨论了子关系的相似性对逻辑推理正确性之间的关系。

### 1 Mereology 理论

Mereology是关于整体和部分关系,整体中部分与部分关

系的理论。在Mereology理论中,部分和整体关系用P表示,定义 $P_{xy}$ 为:概念x是概念y的部分。P具有如下性质:

自反性:概念x是自身的一部分,即 $P_{xx}$ ;

非对称性:概念x和概念y互为部分,则x和y必是同一概念,即 $(P_{xy} \& P_{yx}) \rightarrow x = y$ ;

传递性:若概念x是概念y的部分,概念y是概念z的部分,则概念x是概念z的部分,即 $(P_{xy} \& P_{yz}) \rightarrow P_{xz}$ ;

重叠性:概念x是概念y和概念z的公共部分,则定义概念x是概念y和概念z的重叠,即 $O_{xy} \equiv \exists z(P_{zx} \& P_{zy})$ ;

反重叠性:概念x既包含概念y又包含概念z,则定义概念x是概念y和概念z的反重叠,即 $U_{xy} \equiv \exists z(P_{zx} \& P_{yz})$ ;

原子性:若概念x不再具分成其他部分概念,称x为原子概念,即 $A_x \equiv \neg \exists y(P_{yx})$ ;

弱补充性:概念x是概念y的一部分,则概念y中存在其他的部分概念z,即 $PP_{xy} \rightarrow \exists z(P_{zy} \& \neg O_{zx})$ ;

强补充性:概念x不是概念y的一部分,则概念x中存一部分概念z和概念y没有重叠,即 $\neg P_{xy} \rightarrow \exists z(P_{zx} \& \neg O_{zy})$ ;

一致性:所有概念x的部分都是概念y的部分,反之亦然,则概念x和概念y相同,即 $\forall z(P_{zy} \& P_{zx}) \rightarrow x = y$ ;

和属性:如果概念x和概念y反重叠,则有一概念z,z完全地并仅仅包含了概念x和概念y,即 $U_{xy} \rightarrow \exists z(z = x + y)$ ;

积属性:如果概念x和概念y重叠,则有一概念z为概念x和概念y的重叠部分,即 $O_{xy} \rightarrow \exists z(z = x \cdot y)$ 。

**作者简介:** 连莉(1970—),女,讲师、博士生,主研方向:知识表示,本体重用;马军,博士、教授、博导;雷景生,博士、教授;宋玲,讲师、博士生

**收稿日期:** 2006-04-29 **E-mail:** lianli@sdu.edu.cn

## 2 医学本体中的 Part-Whole 关系的细分和性质分析

Winston 分类法根据 Part-Whole 关系在构造性、同质性和变化性 3 个方面的不同, 将其分成 6 个不同的子类型, 在医学本体中用到了 5 种子类型。

### 2.1 部件/整体(component/integral object)关系

部件/整体关系简称 CO 关系。每个部分都承担了整体的特定功能。在医学本体中, 上层概念主要分为 3 类: 系统, 实体, 实体成员。系统是按照解剖学上功能划分的(如消化系统)。有的系统是器官空间上的结合——消化系统, 有的系统分布全身——内分泌系统。实体就是解剖学上的器官(如胃), 其组成部分有空间的一致性。实体成员是指器官的功能或结构成员。实体成员有的和器官有相同组织类型(肱骨骨节和肱骨), 有的和器官有不同组织类型(肝外胆道和肝)。

CO 关系不满足自反性, 例如, 消化系统自身不是消化系统的一部分。在 CO 关系中, 两个概念之间可能有重叠的部分。例如, 胰既是消化系统的一部分又是内分泌系统的一部分。同样, 两个概念之间可能有反重叠的部分。例如, 胰和肝外胆道都是肝的一部分。

由于 CO 关系不满足自反性, 一个概念必有二个或更多的组成部分。由于部分的功能划分明确, 同一整体的各个部分之间没有重叠, 因此满足弱补充性, 例如, 消化系统中包括胃和与胃没有重叠的肝。由于部分的功能划分明确, 消化系统的器官肝中的各个部分不可能都与内分泌系统重叠, 满足强补充性。

CO 关系不满足一致性。例如肝外胆道由胆囊、胆管、肝管组成, 在图 1 中对肝的描述不同, 图 1(b)中虽然没有肝外胆道, 实际上和图 1(a)指的是同一概念。

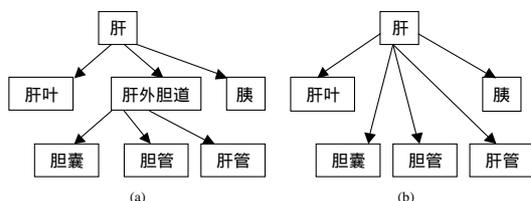


图 1 CO 关系不满足一致性示例

由于部分的功能划分明确, 两个概念交叉的部分必然是一个概念, 满足积属性。反重叠的两个概念(例如肝和胃都是消化系统的部分)不一定有一个和概念(只包括肝和胃), 因此不满足和属性。CO 关系位于医学本体的上层, 每一个概念都有其部分, 所以不满足原子性。

### 2.2 原料/对象(stuff/object)关系

原料/对象关系简称 SO 关系, 表示整体中非独立和不可分割的部分-整体关系。在医学本体中用 SO 关系表达下层概念, 是依据组织胚胎学建立的。

心肌纤维不是自身的原料, 不满足自反性。在 SO 关系中, 两个概念之间可能有重叠的部分。例如, 糖原是骨骼肌纤维的原料又是心肌纤维的原料。同样, 两个概念之间可能有反重叠的部分。例如, 糖原和线粒体都是骨骼肌纤维的原料。由于 SO 关系不满足自反性, 一个概念必有二个或更多的组成部分。由于各个部分不可分割, 同一整体的各个部分之间没有重叠。因此满足弱补充性和强补充性。

由于原料不可再分, SO 关系中的最下层概念具有原子性。由于满足原子性, 图 1 的情况就不可能发生, 满足一致性。骨骼肌纤维与心肌纤维重叠的部分不是一个独立的概念, 因此不满足积属性。也没有只包括线粒体和糖原的概念, 不

满足和属性。

### 2.3 部分/团(portion/mass)关系

部分/团关系简称 PM 关系。PM 关系就是整体的各个部分之间没有明显的界限, 完全是由人任意划分出来的。在医学本体中, 人的体液(例如血液)就是这样一个团, 可以划分为上半身的血液和下半身的血液。划分的方式不同, 造成了 PM 关系不满足一致性。

由于部分划分的任意性, PM 关系满足自反性。团中任意划分的部分之间可能有重叠和反重叠。但是重叠的部分可能不是一个概念, 因此积属性不满足。

### 2.4 地点/区域(place/area)关系

地点/区域关系简称 PA 关系。医学本体中的概念之间还有位置的关系。从人体的横断面——断层图(图 2(a))上可以看出概念之间的区域包含关系。为了论述方便, 首先把断层图简化成图 2(b), 然后由图 2(b)得到树状结构图 2(c)。

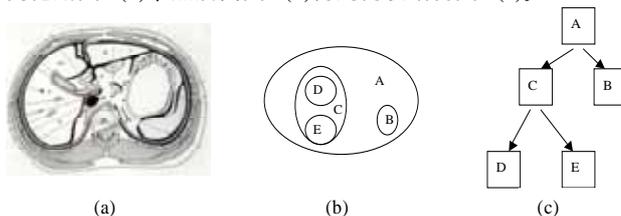


图 2 PA 关系示例

由于 PA 关系中的概念形成了树状结构, 因此不满足自反性, 也不满足重叠性。由于不满足重叠性, 因此对强、弱补充性和积属性的讨论就失去了意义。PA 关系满足反重叠性。例如, 概念 B 和概念 C 的反重叠部分是概念 A。

断层图中给出的概念在空间上是不可再分的, PA 关系满足原子性。由于满足原子性, 图 2 中出现的问题可以避免, 因此 PA 关系满足一致性原理。

### 2.5 成员/团体(member/collection)关系

成员/团体关系简称 MC 关系。在 MC 关系中, 部分是整体的成员(member)。MC 关系中的各个部分之间具有相似的属性或功能。

在医学本体中, 椎骨和脊柱之间就是一种 MC 关系。MC 关系不满足自反性。由于部分的划分明确, MC 关系不满足重叠性。由于不满足重叠性, 对强、弱补充性和积属性的讨论就失去了意义。由于 MC 关系中的部分划分非常确定, 图 1 中出现的问题可以避免, 因此 MC 关系满足一致性。

## 3 种子关系的性质分析

表 1 中用+表示满足, -表示不满足, \*表示讨论无意义。若把各子关系看成长度为 n 的向量。函数 same(a,b)表示向量中对位位置元素的相似性。a, b ∈ {+, -, \*}, 若 a = b, 则 same(a,b)=1, 否则 same(a,b)=0。

表 1 Part-Whole 子关系的性质比较

	自反性	非对称性	传递性	重叠性	反重叠性	弱补充性	补充性	致性	属性	属性	子性
CO	-	+	+	+	+	+					
SO	-	+	+	+	+	+					
PM	+	+	+	+	+	+					
PA	-	+	+	-	+	*					
MC	-	+	+	-	+	*					

定义向量 X, Y 的相似度为

$$Sim(X, Y) = (\sum_{i=1}^n same(X(i), Y(i))) / n$$

得到子关系之间相似度计算结果(表 2)。

表 2 Part-Whole 子关系的相似性

	CO	SO	PM	PA	MC
CO	1	0.727	0.818	0.364	0.364
SO	0.727	1	0.727	0.545	0.455
PM	0.818	0.727	1	0.272	0.272
PA	0.364	0.545	0.272	1	0.909
MC	0.364	0.455	0.272	0.909	1

若定义阈值 $\alpha = 0.7$ ,显然对子关系之间产生了一个划分,即医学本体中的 Part-Whole 关系可分为两类,由集合{CO,SO,PM}和{PA,MC}组成。同集合中的子关系在语义上具有较高的相似性。

因同类子关系的语义是确定的并满足传递性,故使用同一子关系进行三段论推理的结论是正确的。若 X,Y 是属于不同集合中的 Part-Whole 关系的子关系,则采用三段论推理的结论往往是完全错误的。如“手指是乐手的一部分”属于 CO 关系;而“乐手是乐团的一部分”属于 MC 关系,依据三段论推理公式推出“手指是乐团的一部分”显然是错误的。造成错误的主要原因是 CO 和 MC 性质相差太大,属于不同的两集合。

若 X,Y 是属于同一集合的 Part-Whole 的子关系,则采用三段论推理的结论可能是成立的。如,中指是 5 个手指之一(PM);而手指和手掌组成了手(CO),则中指是手的一部分。显然结论是正确的。原因是在同一集合的 Part-Whole 子关系在语义上也非常接近。

作为一般的结论,在 Part-Whole 子关系的基于语义相等的影射中,不在一类的 Part-Whole 子关系之间是不可能建立影射关系的,而影射关系只可能在同类的子关系之间建立。而只有在同一类的 Part-Whole 的子关系之间,有可能进行逻辑推理或基于遗传特征的检索。

## 4 结束语

本文给出对 Part-Whole 关系的细分和对细分后所产生的子关系性质分析的方法学。基于 Winston 分类法,对医学本体建设中常见的 Part-Whole 关系进行了细分,并进一步利用 Mereology 理论对 5 类子关系进行了性质和子关系相似度分析。这种分析不但可帮助理解由本体描述的知识,也可以帮助判断基于多个 Part-Whole 子关系的混合使用的逻辑推理所产生结论的正确性。

但分析并非完美,性质之间并不是完全独立的。如传递性关系可以保证非对称性关系和反重叠性质成立。所以寻找对 Part-Whole 关系分析的更精简的测度标准是未来的研究工作之一。虽然我们的讨论只针对医学本体中的 Part-Whole 关系,但所给出的分析方法也同样适合其他领域。

## 参考文献

- 1 Odell J. Six Different Kinds of Aggregation[M]. Advanced Object-Oriented Analysis & Design Using UML. Cambridge University Press, 139-149.
- 2 Lambrix L P. Padgham Conceptual Modeling in a Document Management Using Part-of Reasoning in Description Logics[J]. Data & Knowledge Engineering, 2000, 32(1): 51-86.
- 3 Rogers J, Rector J A. GALEN's Model of Parts and Wholes: Experience and Comparisons[C]. Proc. of AMIA Symp., 2000: 714-718.
- 4 Mejino J, Agoncillo A, Rickard K, et al. Rosse Representing Complexity in Part-whole Relationships Within the Foundational Model of Anatomy[C]. Proc. of AMIA Symp., 2003: 450-454.
- 5 Ham U, Schulz S, Marko K G. Mereological Semantics for Bio-ontologies [C]. Proc. of AAAI 2004, 2004: 257-262.
- 6 雷玉霞, 睦跃飞, 曹宝香. 基于 Part-Whole 关系的人造物属性分析[J]. 计算机工程与应用, 2004, 40(3): 63-64.

(上接第 70 页)

究分析依然是有待进一步的研究。我们将进一步地研究如何将算法 GFCABD 运用到增量式的数据库中,同时探讨更为有效的方法来减少 GFCABD 的时间复杂度,达到更高的效率。

## 参考文献

- 1 Chen M S. Data Mining: An Overview from a Database Perspective[J]. IEEE Trans. on KDE, 1996, 8(6): 866-883.
- 2 Thuraisingham B. A Primer for Understanding and Applying Data Mining[J]. IT Professional, 2000, 2(1): 28-31.
- 3 Qian Weining, Gong Xueqing, Zhou Aoying. Clustering in Very Large Databases Based on Distance and Density[J]. J. Compt. Sci. & Techno., 2003, 18(1): 67-76.
- 4 Zhang T. BIRCH: An Efficient Data Clustering Method for Very Large Databases[C]. Proc. of the ACM SIGMOD Int'l Conf. on Management of Data. Montreal: ACM Press,1996: 73-84.
- 5 Ester M. A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise[C]. Proc. of the 2<sup>nd</sup> Int'l Conf on Knowledge Discovering in Databases and Data Mining. Portland: AAAI Press,1996.
- 6 Zhang W. Sting: A Statistical Information Grid Approach to Spatial Data Mining[C]. Proc. of the 23<sup>rd</sup> VLDB Conf. Athens: Morgan Kaufmann,1997: 186-195.
- 7 Agrawal R. Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications[C]. Proc. of the ACM SIGMOD Int'l Conf. on Management of Data. Seattle: ACM Press, 1998: 73-84.
- 8 Chen Ning, Chen Anzhou, Long Xiang. An Incremental Grid Density-based Clustering Algorithm[J]. Journal of Software China, 2002, 13(1): 1-7.
- 9 Guan Jihong, Zhou Shuigeng, Bian Fuling, et al. Scaling up the DBSCAN Algorithm for Clustering Large Spatial Databases Based on Sampling Technique[J]. Wuhan University Journal of Natural Sciences, 2001, 6(122): 467-473.
- 10 Dai Weidi, Hou Yuexian, Pi Lianhe, et al. A Clustering Algorithm Based on Building a Density-Tree[C]. Proc. of the 4<sup>th</sup> Int'l Conf. on Machine Learning and Cybernetics, Guangzhou, 2005: 18-21.
- 11 Vijaya, Narasimba Murty P A, Subramanian M, et al. An Efficient Hierarchical Clustering Algorithm for Large Data Sets[J]. Pattern Recognition Letters, 2004, 25(4).