

音乐领域本体的建立和分析^{*}

高颖¹ 曹存根² 睦跃飞²

(中国科学院软件研究所 北京100080)¹ (中国科学院计算技术研究所 北京100080)²

摘要 音乐是一门重要的社会学科,具有丰富的领域知识,这些知识是计算机音乐应用系统所需要的专业知识。本文基于本体论方法形式化地探讨音乐领域本体的建立和分析,具体介绍了本体的建立方法、本体元素的描述、本体的一致性分析等方面内容。我们给出两个判断类间关系一致性的算法并分析其性质,还探讨了基于公理的知识的一致性。

关键词 领域本体,子本体,公理,知识分析,一致性分析

Musical Domain-Specific Ontology Building and Analysis

GAO Ying¹ CAO Cun-Gen² SUI Yue-Fei²

(Institute of Software, Chinese Academy of Sciences, Beijing 100080)¹

(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080)²

Abstract Music is one of important social subjects, and has plenty of domain knowledge, which is necessary for musical computer application system of music. In this paper, a method for musicology-specific ontology building and analysis is introduced. Steps for ontology building, methods of representation for components of ontology, and the consistency of musical ontology are specifically proposed. Two algorithms for checking relationships between categories are presented and their properties are analyzed. Knowledge consistency analysis based on musicology-specific axioms is also discussed.

Keywords Domain-specific ontology, Sub-ontology, Axioms, Knowledge analysis, Consistency analysis

1 引言

人类已经进入信息化社会,正在向知识化社会转变。大量的计算机应用需要学科专业知识的支持,音乐知识系统也不例外(音乐知识查询和音乐智能教学)。即使在一般的自然语言理解中,音乐学知识也是必需的,因为音乐是一般文学作品和新闻报道等文章中经常涉及的内容。“音乐是凭借声波振动而存在、在时间中展现、通过人类的听觉器官而引起各种情绪反应和情感体验的艺术门类^[2]。”首先,音乐学科具有社会科学的一般特点,同时,又由于音乐离不开声音,声音是有物质基础的,从而音乐学科与物理、数学甚至医学都有天然而广泛的联系。总之,在计算机领域,音乐学科知识的表示是十分必要而且具有相当难度的。

近年来,本体论的方法在知识工程领域得到越来越广泛的应用。在很多的有名的知识系统中,如美国 D. Lenat 教授领导研制的大型常识知识库系统 Cyc, Princeton 大学 Berkeley 分校研制的语言知识库 WordNet 等,本体论都有一定的应用。本体论(Ontology)指哲学上的一个学科,20世纪被定义为研究实体存在性和实体存在的本质等方面的通用理论^[13]。一方面,本体论研究深层次上的知识,把知识工程研究中的知识向更深入和更本质的方向推进;另一方面,本体论的研究独立于任何语言,因此本体论将会为不同系统之间知识的共享和互操作提供手段。知识工程中的本体一般指工程上的人造物^[13],较为简单的定义是:本体是被共享的概念化的

一个形式的说明,它是一个形式理论。

知识工程中的知识是泛指,包括不同领域的知识,如军事、医学、民族等等领域。国家知识基础设施 NKI(National Knowledge Infrastructure)旨在建立多领域可共享的知识库,为知识密集的应用,诸如自然语言理解、智能教学等方面提供基础,其基本手段是建立各学科领域本体。由于研究领域不同,各学科间差异很大,因此有必要根据学科特点具体分析各种领域相关问题从而建立相应的领域本体。音乐知识库(Musical Knowledge Base)是 NKI 重要的组成部分,音乐领域本体是建设音乐知识库的重要手段。目前,参照《中国大百科全书》音乐卷的分类体系,经过和专家的多次交流,我们已经建立音乐体裁、作品、乐器、人物几个方面的子本体,并采用人机交互的半自动方法,从《音乐百科词典》(人民音乐出版社,缪天瑞主编)中获取了相当数量的音乐知识。

NKI 已经建立多个学科领域本体,建成了民族、地理、医学、生物、考古等学科知识库。与本课题以往的工作相比,例如王丽丽等人的工作^[23],本文除介绍音乐学科的特点和领域本体的描述方法之外,明确给出了本体的组成,总结了本体的建立步骤,并且提出本体一致性问题,在两个层面上初步分析了本体的一致性。

另一方面,在音乐领域本体的建立和音乐知识的获取方面,国内外的相关工作较少。我们认为,音乐作为一门重要的学科,应该研究它的本体和知识获取。下面,详细介绍音乐领域本体的建立和分析。

^{*} 本文中的工作是在中国科学院计算技术研究所智能信息处理重点实验室知识获取和共享课题组学习期间完成的,本课题得到自然科学基金(#60073017, #60273019)和科技部重大基础研究项目(#2001CCA03000)资助。

2 音乐领域本体概述

2.1 音乐对象

首先,我们确定音乐领域本体的内容和范围。作为一个独立的学科,音乐中的对象十分丰富,对这些对象进行分类的方法也有多种。第一,分为抽象对象和具体对象两类。抽象对象指不占有空间和时间的对象,如音乐体裁、音乐理论等;具体对象指存在于一定的空间或时间中的对象,如乐曲、乐器等。第二,还可以依据对象本身的性质以及在音乐活动中的不同作用分成三大类:音乐作品、音乐规律、音乐活动的参与者。这里的音乐作品是泛指,只要是具有音乐信息的人造物都可算在内,可以是完整的音乐作品,也可以是单独的乐曲片断;音乐规律指人们研究音乐从而形成的认识,包括音乐理论、音乐学、音乐体裁等;音乐活动的参与者指参与演奏演唱过程的对象,包括乐器、演奏艺术、人声、歌唱艺术、人物、团体等。其中,音乐作品是音乐学科中最基本的对象,一切其他的音乐对象都是围绕音乐作品而进行定义和描述的。

上节中列出的两种分类方法结合起来,可以把所有的音乐对象分成如下几大类:作品、理论、音乐学、体裁、乐器、演奏艺术、人声、歌唱艺术、设备、人物、团体。相应地,音乐领域本体包含与上述各类相对应的11个子本体。这11个子本体之间是相对独立的关系,一方面,各子本体自成体系,另一方面,它们并非完全割裂,互相之间有着各种各样的联系。这些子本体有机地组合在一起就形成了完整的音乐领域本体。

2.2 音乐领域本体的组成

音乐领域本体大体由如下5种元素组成:类(categories)、槽(slots)、公理(axioms)、类间关系(relationships)、实例(instances)。

(1)音乐领域本体中的类可称为概念类,是广义上的概念。一般认为,类是具有类似性质的所有个体的抽象,这里的个体称为类的实例。

(2)槽是类和实例的描述符,用以描述对象的一种属性,或与其他对象的关系,并由此分为两类:属性和关系。

(3)公理用于表示一些永真式。在音乐领域本体中,各元素都有特定的含义,元素之间也存在着关联或约束,公理即用来解释这些元素并表示它们之间的关联和约束。

(4)类间关系有多种,其中最基本的是继承关系(is-a),或称“是子类”。通过继承关系,本体中的所有类形成一个有机结合的层次结构。在继承关系基础上还可以定义其他的类间关系。

(5)实例是类包含的基本元素,是本体中最为具体的对象。一个类的所有实例构成此类的外延或称实例集。不同情况下,可以根据本体的描述粒度来决定哪些对象是实例。

下面给出两个类间关系的定义:

定义1 任意两个类 C_1 和 C_2 , 如果任意个体 x , x 是 C_1 的实例蕴含 x 是 C_2 的实例, 那么 C_1 和 C_2 之间存在继承关系(is-a), 记为 C_1 继承 C_2 , 或 is-a(C_1, C_2)。其中, C_2 称为 C_1 的父类(super-category), C_1 称为 C_2 的子类(sub-category)。

由定义知,继承关系是类上的一个偏序。此外,相离也是类之间的一种重要的关系:

定义2 任意两个类 C_1 和 C_2 , 如果不存在个体 x , 使得 x 同时是 C_1 和 C_2 的实例, 那么称 C_1 和 C_2 是相离的(disjoint), 记为 disjoint(C_1, C_2)。注意: disjoint(C_2, C_1)与 disjoint(C_1, C_2)是等价的。

2.3 建立音乐领域本体的原则和方法

前面提到,音乐领域本体包含11个子本体,我们已经建立了音乐体裁、作品、乐器、人物几个子本体。主要有6个步骤:

第一步:合理界定本体的范围或称本体的论域。简单来看,领域本体就是覆盖整个领域的本体,可是具体讨论时常常没有那么简单。一个学科本身的内容往往已经十分丰富,加之学科之间有联系,有时很多内容难于确定是否被该学科所涵盖。音乐领域本体原则上是完全覆盖音乐学科的基本内容,就是说相当于一个比较全面的音乐专业人士所掌握的知识,但参照《音乐百科词典》,我们还加进了许多相关方面的知识,比如音乐团体、音乐设备等。

第二步:确定本体中的类和实例。类和实例有时候并没有严格的界限,很可能一个对象既可以看作类,又可以看作某个类的实例,这时候一方面要注意区分对待,一方面也可以根据本体的描述粒度来决定是作为类还是实例。同时,还要确定每个类的实例集以及每个实例所属的类。

第三步:根据继承关系建立类之间的层次结构。具体的方法有三种:自顶向下法、自底向上法、综合法。自顶向下是先确定领域中一个或几个最一般的类,不断特化现有类,找到新的子类,最终达到最特殊的类;自底向上与自顶向下相反,先确定本体中最特殊的类,不断泛化现有类,最终达到本体中最一般的类;综合法则是上述两种方法的结合。此外,音乐领域本体中还考虑了相离关系,需要在这一步具体明确。

第四步:给各个类定义合适的槽。由于槽规定了类和实例所可能具有的属性和关系,因此槽的定义要保证对类及实例做出完整的描述。此外,由于我们使用框架语言来表示类和实例,因此在定义槽之后,我们还定义了一些侧面对槽做出合理的限制和补充。

第五步:建立公理库。在定义了以上各元素之后,需要深入细致地分析这些元素的含义以及之间的联系和约束,并以一阶语言的形式表述。

第六步:描述类中的实例。同一个类中的实例尽管使用相同的槽,但是槽值各不相同,要根据描述的实例来最终确定槽值。

需要说明的是,领域本体的建立是一个不断重复和修改的过程,没有哪一个时刻的本体是完善的。作为领域知识的一种模型,领域本体需要不断的改进才能接近人对客观世界的认识。

3 音乐领域本体详细描述

3.1 类

音乐领域本体中的类使用 NKI 本体语言^[19]描述。NKI 本体语言是一种框架语言(Frame Language),吸收了 Generic Frame Protocol[Farker et. al., 1998]^[11]的设计思想。类的文法如图1所示。

```

<类定义> ::= defcategory(类名)
           '{'
           <槽定义> { <槽定义> }
           <类自身知识定义> { <类自身知识定义> }
           '}'
<槽定义> ::= <槽类型> : <槽名> | <槽名> : <槽值>
<类自身知识定义> ::= <槽名> : <槽值>
  
```

图1 本体语言描述文法

其中,“槽定义”定义类的子类 and 实例可以使用的槽,一部

分槽有槽值,一部分槽只是给出说明,并不赋值;“类自身知识”是一种特殊的槽,具有槽值,用来定义类本身具有的知识,这部分知识不能被类的子类和实例继承。

本体中类的描述有两种作用:第一,类中包含知识,类中的知识可以是关于类本身的知识,也可以是类实例的共性;第二,类对子类和实例的描述做出规定,限定子类和实例能够使用的描述符(即槽)。

3.2 槽

槽依据性质分为二种类型:

1. 属性槽,简称属性。描述类或实例某一方面的知识;
2. 关系槽,简称关系。相当二元或多元谓词,表示类、实例之间的种种关系。

子类、父类的槽之间有一种继承和发展的关系,情况列举如下:

1. 子类仅继承父类的槽名,并给槽赋新值。如属性“又称”,在不同的类中其值一般不同;
2. 子类继承父类的槽名和部分槽值。如属性“流行地区”,山歌的流行地区显然包括其子类信天游的流行地区;
3. 子类完全继承父类的槽和槽值。如属性“代表曲目”,某个作品类的代表曲目显然也是其父类的代表曲目。

此外,我们定义了一系列侧面对槽进行补充和限制,常用的侧面及含义如表1所示。

表1

侧面	含义
类型	规定槽值取值类型,例如“数量”、“字符串数组”、“时间”等
值域	规定槽值完整的取值范围
不完全值域	规定槽值部分的取值范围
缺省值	槽值的缺省值
下位槽	槽的特殊化或具体化,如属性“使用场合”的下位槽为“演奏场合”、“演唱场合”等
上位槽	槽的一般化,与“下位槽”相对
同义词	槽的同义词
近义词	槽的近义词
逆	槽的逆,如关系“是续篇”与“有续篇”是互逆的
注释	槽的非形式化说明

3.3 类和实例描述举例

为了有一个直观的理解,下面给出音乐作品类(部分)和作品实例《阿黛莱德》的框架表示,分别如图2和图3所示。

defcategory 音乐作品	属性:首演时间
{	:类型 字符串
是子类:作品	属性:首演地点
属性:曲作者	:类型 字符串
:类型 字符串数组	关系:有部分
:同义词 作曲者	:类型 字符串数组
属性:作曲时间	:逆 有部分
:类型 字符串	关系:是部分
属性:作曲地点	:类型 字符串数组
:类型 字符串	:逆 有部分
属性:作品编号	关系:改编为
:类型 字符串数组	:类型 字符串数组
属性:主题调式	关系:改编自
:类型 字符串	:类型 字符串数组
..... }	

图2 音乐作品类(部分)的框架

defframe 《阿黛莱德》:外国声乐作品
{
外文:Adelaide
音乐体裁:歌曲
曲作者:贝多芬
词作者:马蒂松(F. von Matthisson)
作品编号:op. 46
作曲时间:1794年~1795年
内容:该曲抒发了诗人对阿黛莱德的热爱,以及因不得相见而黯然伤神的情感
改编为:李斯特的钢琴曲
}

图3 外国音乐作品《阿黛莱德》的框架

4 音乐领域本体中的公理

4.1 公理的表达方法

我们用一阶语言来表述公理。这里,先对公理中使用的符号做必要的说明。用小写字母 x, y, z 表示类的实例,可以同时指出实例所在的类,例如用“ $\forall x$: 中国音乐作品”来表示 x 是中国音乐作品类的任意一个实例,也可以用谓词的形式“中国音乐作品(x)”来表示。大写字母 X, Y, Z 表示类,用“ $\forall X$: 乐器”表示 X 是任意一种乐器。当需要表示子类或实例的某个槽的取值时,记为 x/X 槽名,如 x 是某一音乐作品,则“ $x \cdot$ 发表时间”表示 x 的发表时间。

为了方便表示,我们还定义了一些谓词和函数,常用的如表2所示。

表2

谓词、函数	含义
和/差(A, B)	函数:求数量 A、B 之和/差
交/并(A, B)	函数:求集合 A、B 之交/并
等于(A, B)	谓词: A、B 之值相等, A、B 的取值可以是数量、集合等等
属于(x, A)	谓词: x 属于集合 A
大于/小于(A, B)	谓词:数量 A 大于/小于数量 B
早于/晚于(A, B)	谓词:时间 A 早于/晚于时间 B
包含于(A, B)	谓词:集合 A 包含于集合 B 中
包含(A, B)	谓词:集合 A 包含集合 B

4.2 槽值公理

槽值公理反映槽值之间的约束和限制,可分为属性间的公理和关系间的公理。

(1) 属性之间的公理

公理1	$\forall x$:中国传统民间器乐乐种(等于($x \cdot$ 演奏乐器, 并($x \cdot$ 主要乐器, $x \cdot$ 其他乐器)))
公理2	$\forall x$:音乐作品(早于($x \cdot$ 作曲时间, $x \cdot$ 发表时间))
公理3	$\forall x$:中国传统民间器乐乐种(等于($x \cdot$ 曲牌数, 和($x \cdot$ 主要曲牌数, $x \cdot$ 其他曲牌数)))
公理4	$\forall x$:中国传统民间器乐乐种(大于等于($x \cdot$ 曲牌数, $x \cdot$ 主要曲牌数))

(2) 关系之间的公理 关系本身具有很多性质,自反性(Reflexivity),传递性(Transitivity),对称性(Symmetry),可逆性(Reversibility)等,我们在寻找关系之间公理的时候,主

要依据关系所具有的性质来考察。

公理5	$\forall x: \text{音乐作品} \forall y: \text{音乐作品} \forall z: \text{音乐作品} (\text{是部分}(x, y) \wedge \text{是部分}(y, z) \rightarrow \text{是部分}(x, z))$
公理6	$\forall x: \text{音乐作品} \forall y: \text{音乐作品} \forall z: \text{音乐作品} (\text{有部分}(x, y) \wedge \text{有部分}(y, z) \rightarrow \text{有部分}(x, z))$
公理7	$\forall x: \text{音乐作品} \forall y: \text{音乐作品} (\text{相似于}(y, x) \rightarrow \text{相似于}(x, y))$
公理8	$\forall x: \text{音乐作品} \forall y: \text{音乐作品} (\text{是续篇}(y, x) \rightarrow \text{有续篇}(x, y))$

4.3 内涵公理

内涵公理的内容很广泛,主要用来规定本体中其他元素的含义。较上述两类公理,这类公理往往需要深入地考察才可以发现。下面列举音乐领域本体中的部分内涵公理。

公理9	$\forall X: \text{西方音乐} \forall y: \text{音乐作品} (\text{等于}(y, X \cdot \text{第一部作品}) \rightarrow \text{等于}(X \cdot \text{最早作曲家}, y \cdot \text{作曲者}))$
公理10	$\forall x: \text{音乐作品} \forall Y: \text{舞曲} (\text{是实例}(x, Y) \rightarrow \text{属于}(x \cdot \text{节拍}, Y \cdot \text{节拍}))$
公理11	$\forall x: \text{音乐作品} \forall y: \text{音乐作品} (\text{等于}(x \cdot \text{曲作者}, y \cdot \text{曲作者}) \wedge \text{等于}(x \cdot \text{作品编号}, y \cdot \text{作品编号}) \rightarrow \text{等于}(x, y))$
公理12	$\forall x: \text{音乐作品} (\text{集成剧}(x) \vee \text{集成曲}(x) \rightarrow \text{大于}(1, x \cdot \text{曲作者数}))$
公理13	$\forall x: \text{音乐作品} (\text{三部曲}(x) \vee \text{三部曲}(x) \rightarrow \exists y_1, y_2, y_3: \text{音乐作品} (\text{是成员}(y_1, x) \wedge \text{是成员}(y_2, x) \wedge \text{是成员}(y_3, x)))$
公理14	$\forall x: \text{音乐作品} (\text{套曲}(x) \rightarrow \text{大于}(x \cdot \text{包含乐曲数}, 1))$
公理15	$\forall x: \text{中国传统民间器乐} \forall Y: \text{乐器} (\text{属于}(Y, x \cdot \text{演奏乐器}) \rightarrow \text{是子类}(Y, \text{中国民族乐器}))$
公理16	$\forall x: \text{弦索类乐种} \forall Y: \text{乐器} (\text{属于}(Y, x \cdot \text{演奏乐器}) \rightarrow \text{是子类}(Y, \text{弦乐器}))$
公理17	$\forall x: \text{丝竹类乐种} \forall Y: \text{乐器} (\text{属于}(Y, x \cdot \text{演奏乐器}) \rightarrow \text{是子类}(Y, \text{弦乐器}) \vee \text{是子类}(Y, \text{管乐器}))$
公理18	$\forall x: \text{鼓吹类乐种} \exists Y: \text{乐器} (\text{是子类}(Y, \text{吹奏乐器}) \wedge \text{属于}(Y, x \cdot \text{主要乐器}))$
公理19	$\forall x: \text{吹打类乐种} \forall Y: \text{乐器} (\text{属于}(Y, x \cdot \text{主要乐器}) \rightarrow \text{是子类}(Y, \text{管乐器}) \vee \text{是子类}(Y, \text{打击乐器}))$
公理20	$\forall x: \text{锣鼓类乐种} \forall Y: \text{乐器} (\text{属于}(Y, x \cdot \text{演奏乐器}) \rightarrow \text{是子类}(Y, \text{打击乐器}))$

5 音乐领域本体的一致性分析

简单说,本体的一致性指本体包含的所有知识之间没有矛盾,其各组成部分构成一个协调的整体。下面分两个层次介绍这方面的工作:类间关系的一致性,基于公理的知识一致性。

5.1 类间关系的一致性分析

音乐领域本体中的类数目很大,各类间的关系十分复杂,诸如按体裁划分的各作品类、各乐器类等,这些类往往不能组织成一个规整的结构,因此容易产生类间关系不一致的现象。本节具体讨论两种不一致。

首先,由前面的定义1和定义2有如下推论:

推论1 本体中的任意类 C_1 , 有 $\text{is-a}(C_1, C_1)$ 。

推论2 本体中的任意两个互不相同的类 C_1, C_2 , 如果有 $\text{is-a}(C_1, C_2)$, 则 $\text{is-a}(C_2, C_1)$ 不成立。

推论3 本体中的任意类 C_1, C_2, C_3 , 如果有 $\text{is-a}(C_1, C_2), \text{is-a}(C_2, C_3)$, 则有 $\text{is-a}(C_1, C_3)$ 。

推论4 本体中的任意类 C_1, C_2, C_3, C_4 , 如果有 $\text{is-a}(C_1, C_3), \text{is-a}(C_2, C_4)$, 并且 $\text{disjoint}(C_3, C_4)$, 则有 $\text{disjoint}(C_1, C_2)$ 。

类间的继承和相离关系中,容易出现两个不一致:循环继承和关系冲突。循环继承指多个类之间的继承关系形成一个环状;关系冲突指两个类之间同时具有继承和相离的关系。注意:由于目前我们在本体中只定义了继承和相离这两种类间关系,因此本文只讨论这两种关系的冲突。如果用图4来表示,本体中不能出现下面的两种情况。

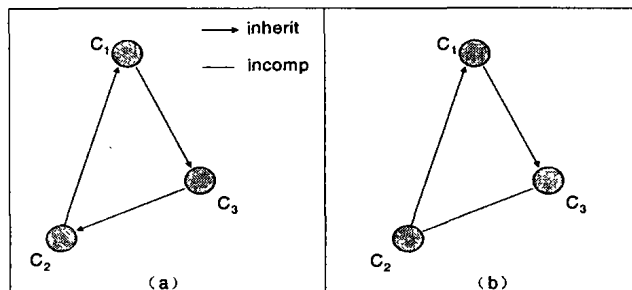


图4 循环继承(a)和关系冲突(b)

图4表示了循环继承和关系冲突的情况。(a)显示三个不同的类 C_1, C_2, C_3 互相继承,在我们的本体中是不允许出现的,因为这表明三个类的实例集相同,三个类是同一个类;(b)显示 C_2 是 C_3 的子类,但 C_2 和 C_3 又相离,依据继承和相离关系的定义,这是矛盾的,所以这种情况在本体中也不允许出现。下面给出两个算法分别检查这两种不一致。

算法1 检查循环继承的算法

1. 设本体中存在 n 个互不相同的类 $C_1, \dots, C_n, n \geq 1$;
2. $i = 1$;
3. $S_1 = \phi, S_2 = \phi$;
4. 对本体中的所有类 C' , 如果 $\text{is-a}(C_i, C')$, 则 $S_1 = S_1 \cup Y\{C'\}$; 如果 $\text{is-a}(C', C_i)$, 则 $S_2 = S_2 \cup Y\{C'\}$;
5. $S_1 = S_1 \setminus \{C_i\}, S_2 = S_2 \setminus \{C_i\}$;
6. 如果 $S_1 \cap S_2 \neq \phi$, 算法结束,则本体中存在循环继承;
7. $i = i + 1$;
8. 如果 $i \leq n$, 转3;
9. 算法结束,本体中不存在循环继承。 □

直观上,算法1的判断依据是:设类 C 的所有父类构成的集合(不含 C)为 S_1 , 所有子类构成的集合(不含 C)为 S_2 , 如果 S_1 和 S_2 具有公共的元素,则表示出现循环继承。

算法2 检查关系冲突的算法

1. 设本体中存在 n 个互不相同的类 $C_1, \dots, C_n, n \geq 1$;
2. $i = 1$;
3. $S_1 = \phi, S_2 = \phi$;
4. 对本体中的所有类 C' , 如果 $\text{is-a}(C_i, C')$ 或者 $\text{is-a}(C', C_i)$, 则 $S_1 = S_1 \cup Y\{C'\}$; 如果 $\text{disjoint}(C_i, C')$, 则 $S_2 = S_2 \cup Y\{C'\}$;
5. 如果 $S_1 \cap S_2 \neq \phi$, 算法结束,则本体中存在关系冲突;
6. $i = i + 1$;
7. 如果 $i \leq n$, 转3;
8. 算法结束,本体中不存在关系冲突。 □

直观上,算法2的判断依据是:设类 C 的所有父类和所有子类构成的集合为 S_1 , 所有与 C 相离的类构成的集合(不含 C)为 S_2 , 如果 S_1 和 S_2 具有公共的元素,则表示出现关系冲突。

下面分析算法的可靠性和完备性。

如果类间关系是循环继承的,那么由算法1一定能够得到类间关系是循环继承的,这是算法1的完备性。如果算法1得到类间关系是循环继承的,那么类间关系一定是循环继承的,这是算法1的可靠性。相应地有算法2的完备性和可靠性。下面给出几个定理,证明省略。

定理1 算法1和算法2一定会终止。

定理2 算法1和算法2是完备的。

定理3 算法1和算法2是可靠的。

5.2 基于音乐公理的知识一致性分析

公理是学科本体中重要的组成部分,公理的作用有二:第一,公理即知识;第二,公理保证本体中较低层次的知识是一致的。这里所说的知识一致性指,在公理约束下的知识是与公理没有矛盾的。由知识工程师从文本中获取的知识不总是正确的和没有矛盾的,知识来源的不准确,知识获取过程中人为的疏忽等等原因,都可能导致知识是不一致的。这里给出几个例子来说明。

例1:

$\forall x$: 中国传统民间器乐乐种(等于(交(x ·演奏乐器, x ·排斥乐器), \emptyset))

上述公理说明对于任意中国传统民间器乐乐种,其演奏乐器和排斥乐器的值相交为空,即不可能有同一种乐器,既为此乐种所使用,又为其所排斥。如果知识库中有如下知识,即可被及时地发现并纠正:

```
defframe 乐种1:中国传统民间器乐乐种
{
  演奏乐器:锣,和鼓,和琵琶等
  排斥乐器:锣
  .....
}
```

例2:

$\forall x$: 鼓吹乐类乐种 $\exists Y$: 乐器(是子类(Y , 吹奏乐器) \wedge 属于(Y , x ·主奏乐器))

上述公理的含义是,任意鼓吹乐类乐种都以某一种或几种吹奏乐器为其主奏乐器,这是鼓吹乐类乐种的内涵之一。显然下述知识与公理矛盾,因为中国传统民族乐器二胡和阮是弦乐器,而非吹奏乐器:

```
defframe 乐种2:鼓吹乐类乐种
{
  主奏乐器:二胡,和阮
  .....
}
```

应该特别指出的是,本体的一致性分析是一件十分复杂的事情,涉及类、实例、公理各个层次的知识,知识工程师需要统一考虑,并找出有效的方法来检验其一致性。上述的讨论只是一些初始的工作,以后可以尝试从局部分析,在保证本体局部一致性的前提下分析整个本体的一致性。

总结及今后的工作 本文基于本体理论提出一种音乐学科知识的表示方法。类、槽、类间关系、公理、实例是本体的基本组成部分,我们具体探讨了建立本体的主要方法,定义了本体描述语言,并给出类和实例框架的例子。在此基础上,初步分析本体的一致性,讨论了类间关系以及基于公理的知识的一致性。目前,按照这一方法,我们已经从《音乐百科词典》中获取了相当数量的学科知识,完成了几个领域子本体的建设。但是,仍有大量的工作需要完成,相当数目的问题也急需解决。首先,音乐学科中与物理、数学、医学等学科关系紧密的部分,诸如理论、人声等还没有进行全面处理,这部分知识极为复杂而且十分严谨,需要在深入掌握其中规律的基础上合理建立本体;第二,已有子本体的公理库还不完整,找全子本体

的公理是一件十分困难的事情,为了保证本体是一致和完备的,要尽可能地完善公理库;第三,本体的一致性要求更有效的方法来检验,本文所介绍的类间关系的一致性判断算法仅仅是这一部分工作的开始,更加深入地探讨要在今后完成;第四,基于本体进行文本知识自动获取是本体的重要应用,我们已经实现了部分文本知识的半自动获取,今后希望在较为完善的本体的支持下,能够实现更进一步的知识自动获取。

致谢 本文的工作得到北京师范大学华进老师的支持,在此表示感谢!

参考文献

- 1 缪天瑞. 音乐百科词典. 人民音乐出版社, 1998
- 2 中国大百科全书——音乐卷. 中国大百科全书出版社
- 3 袁静芳. 中国传统音乐概论. 上海音乐出版社, 2000
- 4 于润洋. 西方音乐通史. 上海音乐出版社, 2001
- 5 周青青. 中国民歌. 人民音乐出版社, 1993
- 6 Welty C. The Ontological Nature of Subject Taxonomies. In: Proc. of the First Intl. Conf. (FOIS'98), Trento, Italy. 317~327
- 7 Guarino N. Formal Ontology and Information Systems. In: Proc. of the First Intl. Conf. (FOIS'98), Trento, Italy
- 8 Guarino N, Welty C. Ontological Analysis of Taxonomic Relationship. In: The Intl. Conf. on Conceptual Modeling, Oct. 2000
- 9 Guarino N, Welty C. A Formal Ontology of Properties. In: Proc. of 12th Intl. Conf. on Knowledge Engineering and Knowledge Management, Lecture Notes on Computer Science, Springer Verlag, 2000
- 10 Smith B, Mark D M. Ontology and Geographic Kinds. In: Proc. of the Intl. Symposium on Spatial Data Handling (SDH'98), Vancouver, Canada, 1998
- 11 Chaudhri V K, Farquhar A, et al. The Generic Frame Protocol 2.0: [SRI International Technical Report]. 1997
- 12 Noy N F, McGuinness D L. Ontology Development 101: A Guide to Creating Your First Ontology
- 13 陆汝钊. 世纪之交的知识工程与知识科学. 清华大学出版社, 2001
- 14 石纯一, 黄昌宁, 王家庆. 人工智能原理. 清华大学出版社, 1993
- 15 王永庆. 人工智能原理与方法. 西安交通大学出版社, 1998
- 16 陆汝钊. 人工智能. 科学出版社, 1995
- 17 陆钟万. 面向计算机科学的数理逻辑. 科学出版社, 1998
- 18 曹存根. 面向专家的知识获取. 科学出版社, 1998
- 19 Cao Cungen, Feng Qiangze, Gao Ying, et al. Progress in the Development of National Knowledge Infrastructure. Journal of Computer Science and Technology, 2002, 17: 523~534
- 20 Gu Fang, Cao Cungen, Sui Yuefei. A Domain-Specific Ontology of Botany, to appear in Journal of Computer Science and Technology
- 21 Tian Wen, Gu Fang, Cao Cungen. Designing a Top-Level Ontology of Human Beings: A Multi-Perspective Approach. Journal of Computer Science and Technology, 2002, 17: 636~656
- 22 张德海, 曹存根, 张宇翔. 国家和城市知识获取与本体论分析. 见: 中国人工智能学会第九届全国学术年会暨中国人工智能学会成立20周年庆祝大会, 2001. 366~370
- 23 王丽丽, 曹存根, 顾芳, 田雯. 基于本体论的民族知识获取与分析. 计算机科学, 2003, 30(5): 47~54, 81