



(12)发明专利

(10)授权公告号 CN 104199882 B

(45)授权公告日 2018.04.10

(21)申请号 201410418075.1

(22)申请日 2014.08.22

(65)同一申请的已公布的文献号
申请公布号 CN 104199882 A

(43)申请公布日 2014.12.10

(73)专利权人 北京航空航天大学
地址 100191 北京市海淀区学院路37号

(72)发明人 张霖 胡安瑞 胡晓航 李博文
麦艺帆 陶飞

(74)专利代理机构 北京科迪生专利代理有限责
任公司 11251

代理人 杨学明 顾炜

(51)Int. Cl.
G06F 17/30(2006.01)

(56)对比文件

CN 1647070 A, 2005.07.27,
CN 102968450 A, 2013.03.13,
WO 0217097 A1, 2002.02.28,

王英林等.基于本体的可重构知识管理平台.《计算机集成制造系统》.2003,第9卷(第12期),第1136-1144页.

张晓蕾.基于OWL的语义查询技术研究.《中国优秀硕士学位论文全文数据库 信息科技辑》.2009,(第9期),第1138-881页.

张景雨等.基于模板的知识集成研究.《微电子学与计算机》.2009,第26卷(第4期),第203-205、209页.

审查员 尤鑫

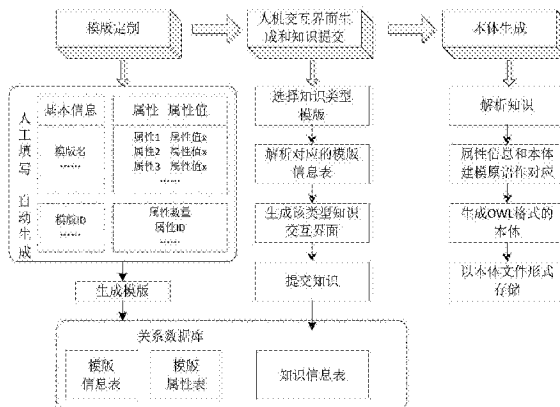
权利要求书2页 说明书6页 附图3页

(54)发明名称

一种基于智能模版定制的结构化知识及其本体的获取方法

(57)摘要

本发明提供一种基于智能模版定制的结构化知识及其本体的获取方法,该方法首先提出了模版定制的思想,为领域专家提供针对结构化知识获取的模版定制工具.领域专家可以根据该工具定制其领域内某类型结构化知识的基本信息和该知识的属性以及属性值.系统根据定制好的模版自动生成该类型结构化知识获取的人机交互界面及其数据库存储表.知识提供者可以通过该界面提交知识.系统将这些知识存储到的关系数据库中,并自动生成对应的OWL格式的本体文件.在结构化知识获取过程中,该方法的灵活度大,节省资源,摆脱了传统上需要事先制作好模版的依赖.同时,模版定制者不需要计算机编程能力.本方法是一种高效、快速、方便的结构化知识获取方法.



1. 一种基于智能模版定制的结构化知识及其本体的获取方法,其特征在于,该方法包括3个步骤:

步骤1)、模版定制;

该模版定制工具可以为领域专家提供一个人机交互界面用来定制模版,领域专家可以为各类结构化知识设计专有的获取模版;

步骤1)提出了模版定制思想,具体实现方式如下:

通过人机交互的方式,领域专家可以定制该类知识的基本信息以及知识的属性和属性值,定制完成后系统会把这些信息以表的形式存储在关系数据库中;在整个过程中一共有三类表,分别是模版信息表、模版属性表和知识信息表;每一个知识类型在模版信息表中有一条记录,并对应一张关于它的属性和属性值的模版属性表,以及基于知识的属性和属性值生成的知识信息表;模版信息和具体的知识都存储在这三类表中;

步骤2)、人机交互界面生成和知识提交;

解析模版的具体内容,生成该类结构化知识获取的人机交互界面,知识提供者可以使用该界面提交知识;

步骤2)在步骤1)的基础上实现人机交互界面生成和知识提交,具体实现方式如下:

知识提供者选择要提供的知识类型,系统会找到该类型知识对应的模版,在关系数据库中解析模版信息表中关于该模版的记录及其对应的模版属性表,生成一个该类知识的人机交互获取界面,所有需要填写的信息都会显示出来供填写,知识提供者可以通过填写该界面提交知识,所有通过该模版生成的界面提交的知识都会被插入到对应类型的知识信息表中,系统可以解析提交过的内容供用户查看;

步骤3)、本体生成;

用户提交的知识会根据该类知识的信息,自动生成OWL格式的本体描述文件;

步骤3)在步骤2)的基础上解析用户提交的知识,以本体的方式对知识建模并输出OWL格式本体文件,具体实现方式如下:

系统提取数据库中关于某条知识的所有信息进行解析,将这些信息和相关建模原语相对应,即转化为类(Class)关系、数据属性(Datatype Property)和关系属性(Object Property),进而,系统通过Jena工具包,以OWL为本体建模标准对知识进行建模,以本体文件的形式存储;

该基于智能模版定制的结构化知识及其本体的获取方法灵活高效地结构化知识获取,本方法中,可以根据知识类型,随时设计对应的知识获取模版并生成人机交互界面,结构化的知识类型多的问题被解决;

该基于智能模版定制的结构化知识及其本体的获取方法合理利用了系统空间,传统方法中,尽量多的为可能的结构化知识设计模版,可能有的知识有了模版而没有知识,而在本方法中,只有确定了系统中没有对应的知识获取模版后,可立刻为新知识类型设计模版,这样就节省了系统空间;

该基于智能模版定制的结构化知识及其本体的获取方法随时修改结构化知识的属性,本方法中,可以通过修改知识模版而使获取界面和后台数据库结构得到修改,从而实现已有知识类型的修改;

该基于智能模版定制的结构化知识及其本体的获取方法快速高效的获取结构化知识的本

体获取,可以在知识提交后自动生成知识的本体文件。

一种基于智能模版定制的结构化知识及其本体的获取方法

技术领域

[0001] 本发明设计一种基于智能模版定制的结构化知识及其本体的获取方法,即利用一种可定制的模版,为知识提供者提供结构化知识提交的界面。实现不同类型的结构化知识的获取。从而解决由于多领域的结构化知识种类繁多造成的获取繁琐的问题。同时提供了一种本体获取方法,实现结构化知识本体快速获取,属于计算机领域。

背景技术

[0002] 进入21世纪以来,随着信息技术、计算机技术、互联网技术的发展,知识扮演着愈发重要的角色。在计算机领域,数据、信息和知识是研究的重点和热点。知识是对数据和信息的抽取和深化,是对事物更深层次的认识。以知识为研究点的知识工程作为由人工智能发展而来的科学,重点研究知识表示、知识获取和知识应用等问题。其目的是将存储在文本、图像、声音及思想等中的人类知识,抽取出来并表示成计算机可以识别的形式,存储起来。开发相应的处理软件,准确地为用户提供其需要的知识,或合理的利用知识进行推理来解答一些问题。借助于计算机强大的计算能力,极大地提高用户工作的效率。

[0003] 由于知识的形式多种多样,有些知识是模糊的或抽象的,即使是人类语言也很难描述,表示成计算机可以识别的形态就更加困难。同时,知识的载体也不同,针对有些载体内知识的抽取很难实现。所以在知识工程中,知识获取一直是研究重点和难点。目前常见的知识获取方式可以分成人工获取、自动获取和半自动获取三类。人工获取是由知识工程师以阅读或与某领域专家交流的方式,获取该领域内的知识,然后将这些知识编写成计算机可以识别的形式的一种获取方式。这种方式的获取效率较低。自动获取是通过编写好的计算机程序自动在知识载体中抽取知识的方式获取知识,这种方式目前只能在极小的特殊范围内实现,还是一种比较理想化的知识获取方式。目前最常见的知识获取方式是半自动知识获取方式。知识工程师首先获取某领域知识的关键属性并设计知识获取模版。该模版包括人机交互界面和对应的关系数据库表。知识提供者通过人机交互界面填写知识,系统自动将其存储到数据库中设计好的表中。这种方法尤其适合结构化知识的获取。以计算机CPU为例,主要属性有基本参数、频率、插槽、内核、缓存、技术参数等,知识提供者可以填写这些属性,实现不同类型CPU的知识获取。这是目前使用最广的知识获取方法。但是该方法的最大问题是结构化知识的种类太多,不同的事物的属性各不相同。每一个模版都要设计一个交互界面及其对应的数据库表。制作全部的知识获取模版是不现实的。再有,制作这样的模版需要计算机编程方面的能力,而领域专家和知识提供者并不都具备这样的能力。同时,知识表示形式有很多。其中,本体作为目前较为流行的表示形式,特别适合概念及其关系一类知识的表示。目前的本体获取往往使用软件进行人工编写。如使用Protégé编写并生成本体表示的OWL文件。这种方法在结构化知识本体获取上效率较低。

[0004] 针对目前结构化知识获取中种类繁多和本体获取繁琐的问题,本发明设计了一种智能模版定制的结构化知识获取方法。该方法可以由领域专家自行设计知识获取模版,由设计好的模版生成对应的人机交互界面及对应的数据库存储表,供知识提供者提交相应的

知识并存储。同时生成知识的本体OWL文件。

[0005] 本发明目的为:结构化知识的获取中,知识种类繁多,为每一类知识都设计获取模版是不现实的。同时目前流行的本体获取方式较为繁琐,无法快速高效地实现结构化知识的本体表示。本发明设计了一种智能模版定制的方法。通过该方法,可以由该领域的专家自行设计结构化知识的获取模版,并由这些模版生成对应的知识提交界面和数据库表,知识提供者通过填写这些界面提交知识,系统将其存入数据库中,并可以自动生成这些结构化知识的本体表示文件。

[0006] 目前对于结构化知识的获取一般采用人机交互获取方式。以CPU为例,其重要属性有CPU的型号、基本参数、频率、插槽、内核、缓存、技术参数等,具体属性包括基本参数(适用类型、所属系列、包装尺寸),频率(主频、最大睿频、外频、倍频),插槽(插槽类型、针脚数目),内核(核心代号、核心数量、线程数、制作工艺、热设计功耗),缓存(二级缓存、三级缓存),技术参数(指令集、内存控制器、最大支持内存、超线程技术、虚拟化基础、处理器位数)等。CPU的描述知识是一种典型的结构化知识,这里知识的获取一般是根据上述的信息开发一个人机交互页面。该页面包含了上述的这些属性,并空出属性的值供知识提供者填写。用户依次填写了这些值并提交。系统获取到这些值后存储到后台提前设计好的关系数据库中。这样就实现了一个CPU的知识获取。

[0007] 在这个过程中存在很多问题。首先,知识获取模版的不确定性,如果系统中没有CPU的知识获取模版,则知识提供者无法提交有关的知识。另外,获取模版的柔性不足,则会造成知识获取不准确。如有的CPU是附带显卡的。如果上述的模版中没有对应的属性,则这部分的知识无法获取。同时,这类非结构化的知识类型太多,如内存、齿轮、机床等,甚至组织信息、公司结构等。每一个类型的知识都设计一个获取模版是不现实的。而无论是设计还是修改这样的模版需要一定的计算机编程能力,领域专家和知识提供者在没有系统工程师的帮助下无法实现。再有就是这类知识的本体获取太过繁琐,传统方法是使用软件人工编写。如果能够自动生成本体表示文件,则会极大提高结构化知识的本体获取效率。

发明内容

[0008] 本发明为解决上述问题,提出了一种基于智能模版定制的结构化知识及其本体的获取方法。该方法首先开发一个模版定制工具,领域专家可以使用该工具建立或修改知识获取模版。建立或修改后的模版存储在数据库中。然后开发一个模版生成工具,该工具可以解析数据库中的模版信息,生成对应的人机交互界面及其数据库存储表。知识提供者可以通过填写该界面提交知识并存入数据库。知识提交后,系统通过本体生成工具自动生成OWL格式的本体文件存入系统硬盘。

[0009] 本发明采用的技术方案为:一种基于智能模版定制的结构化知识及其本体的获取方法,该方法包括3个步骤:

[0010] 步骤1)、模版定制。

[0011] 该模版定制工具可以为领域专家提供一个人机交互界面用来定制模版。领域专家可以为各类结构化知识设计专有的获取模版。

[0012] 步骤2)、人机交互界面生成和知识提交。

[0013] 解析模版的具体信息,生成该类结构化知识获取的人机交互界面。知识提供者可

以根据该界面提交知识。

[0014] 步骤3)、本体生成。

[0015] 用户提交的知识会根据该类知识的信息,自动生成OWL格式的本体描述文件。

[0016] 进一步的,步骤1)提出了模版定制思想,具体实现方式如下:

[0017] 通过人机交互的方式,领域专家可以定制该类知识的基本信息以及知识的属性和属性值。定制完成后系统会把这些信息以表的形式存储在关系数据库中。在整个过程中一共有三类表,分别是模版信息表、模版属性表和知识信息表。每一个知识类型在模版信息表中有一条记录,并对应一张关于它的属性和属性值的模版属性表,以及由知识属性和属性值生成的知识信息表。模版信息和具体的知识都存储在这三类表中。

[0018] 进一步的,步骤2)在步骤1)的基础上实现人机交互界面生成和知识提交,具体实现方式如下:

[0019] 知识提供者选择要提供的知识类型,系统会找到该类型知识对应的模版,在关系数据库中解析模版信息表中关于该模版的记录及其对应的模版属性表。生成一个该类知识的人机交互获取界面。所有需要填写的属性都会显示出来供填写。知识提供者可以通过填写该界面提交知识。所有通过该模版生成的界面提交的知识都会被插入到对应类型的知识信息表中。系统可以解析提交过的内容供用户查看。

[0020] 进一步的,步骤3)在步骤2)的基础上解析用户提交的知识,以本体的方式对知识建模并输出OWL格式本体文件,具体实现方式如下:

[0021] 系统提取知识信息表中描述该知识的基本信息和模版属性表中的属性信息和属性值信息进行解析,将这些信息和相关建模原语相对应,即转化为类(Class)关系、数据属性(Datatype Property)和关系属性(Object Property)。进而,系统通过Jena工具包,以OWL为本体建模标准对知识进行建模,以本体文件的形式存储。

[0022] 本发明与现有技术相比的优点为:

[0023] 1. 灵活高效地结构化知识获取。而本方法中,可以根据知识类型,随时设计对应的知识获取模版并生成人机交互界面。结构化的知识类型多的问题被解决。

[0024] 2. 合理利用了系统空间。传统方法中,尽量多的为可能的结构化知识设计模版,可能有的知识有了模版而没有知识。而在本发明中,只有确定了系统中没有对应的知识获取模版后,可立刻为新知识类型设计模版。这样就节省了系统空间。

[0025] 3. 随时修改结构化知识的属性。本发明中,可以通过修改知识模版而使获取界面和后台数据库结构得到修改,从而实现已有知识类型的修改。

[0026] 4. 快速高效的结构化知识的本体获取。可以在知识提交后自动生成知识的本体文件。

附图说明

[0027] 图1为本方法整体示意图;

[0028] 图2为模版定制工具示例;

[0029] 图3为模版定制过程示例;

[0030] 图4为模版存储关系表结构;

[0031] 图5为生成的知识提交界面示例;

[0032] 图6为示例的本体中建模原语的对应关系。

具体实施方式

[0033] 下面结合附图对本发明作进一步详细的描述。

[0034] 本发明是一种基于智能模版定制的结构化知识及其本体的获取方法,如图1所示,主要包括三个部分:模版定制,人机交互界面生成和知识提交,本体生成。

[0035] (1) 模版定制

[0036] 定制工具提供一个模版的定制界面:

[0037] 通过该界面可以填写知识的名称,并定义知识的属性数量、属性名称以及属性值。属性的数量是通过每次点击添加属性扩展得到的。属性的名称由用户自定义。属性值由属性值的类型和对应类型的内容组成。属性值的类型由系统事先定义好,常见的类型如纯文本(填写内容)、选择框(单选框或多选框)、下拉菜单(多种类的单选)等。对应类型的内容定义了该属性值类型中可以填写的范围,如下拉菜单中可选的条目等。用户选择类型并填写该类型的内容。在模版说明中附填写说明,详细介绍不同类型属性值的内容填写方法。该工具的具体示例如图2所示。具体的填写过程的示例如图3所示。

[0038] 模版定制过程一般由领域专家来实施。因为模版定制需要对该领域有深层次的了解。可以通过设置系统角色实现相应的权限控制,来保证模版的正确性和可用性。领域专家填写完成后,点击定制新模版。对应的模版名称、模版ID(唯一编号)、属性数量、属性名称、属性ID(唯一编号)、属性值等,一并存入关系数据库中描述模版的数据表中。此时一个结构化知识的模版定制完成。

[0039] 系统中主要包含三类表,其结构关系如图4所示。第一类表是模版信息表,该表中存储了系统所有提交的模版的基本信息,如模版名称,模版ID,模版制作用户等,结构如表1所示。该表中每一条记录一个知识模板的基本信息。第二类表是模版属性表,结构如表2所示,该表根据领域专家提交的属性和属性值自动生成,记录了该类知识具有的属性和属性值。具体示例如表3所示。第三类表是知识信息表,其结构如表4所示。该表由表2的属性表生成,其中每个属性为该表的字段,属性值即为字段的定义。每一条记录表示某类型的具体的每一条知识的内容,具体示例如表5所示。另外,可以通过点击查看已有模版对现有模版进行修改,点击提交后会自动修改后台的关系数据库表格。每一个类型的知识对应表1的一条记录和一张表2以及一张表4。

[0040] 表1 模板信息表

模板信息表(Model)					
	数据段名	类型	长度	说明	
[0041]	1	Modelid	varchar	20	数据表唯一编号,主键 1
	2	Modelname	varchar	30	模板名称
	3	Username	varchar	20	用户名

[0042] 表2 模板属性表

[0043]

模板属性表(Modelname+Username)				
	数据段名	类型	长度	说明
1	Propertyid	varchar	20	数据表唯一编号,主键 1
2	Propertyname	varchar	30	属性名称, 外键
3	Propertytype	varchar	20	属性类型
4	Propertycontent	varchar	20	属性内容

[0044] 表3 模板属性表示例

[0045]

Propertyid	Propertyname	Propertytype	Propertycontent
pid1	Name	文本显示	
pid2	Gende	选项显示(单选)	
pid3	Age	选项显示(下拉型单选)	10~20 20~30 30~40 >40

[0046] 表4 知识信息表

[0047]

知识信息表(Modelname+Username)				
	数据段名	类型	长度	说明
1	Propertyname1	varchar	20	属性 1
2	Propertyname2	varchar	20	属性 2
3	Propertyname3	varchar	20	属性 3

[0048] 表5 知识信息表示例

[0049]

Name	Gender	Age
zhangsan	Male	20~30
lisi	Male	30~40

[0050] (2) 人机交互界面生成和知识提交

[0051] 用户选择要提交的知识类型,系统根据选择解析关系数据库中该类型知识的模版。由表1中关于该知识的那一条记录,以及该记录对应的表2和表4的内容,生成知识获取的交互界面,具体示例如图5所示。

[0052] 在这个过程中,系统首先读取模版信息表中该条模版的记录。查询到该模版的名字,ID和定制者。根据模版的ID找到对应该模版的模版属性表。查询该表中的属性和属性值等。根据上面两部的查询结果,将模版的名字、定制者等信息显示在模版显眼处。属性以列表的方式显示在页面中间。每一行前面显示属性的名称,后面根据属性值的类型及其内容生成一个填写框。如文本类型就产生一个文本框,选择类型就产生选择框,下拉类型产生一个下拉菜单。知识提交者通过填写这些填写框来提交知识。知识提供者提交知识以后,会将内容插入到知识信息表的中。这时就完成了结构化知识的提交。所有通过该模版生成的界

面而提交的知识都被存储在这个表中。系统可以解析这个表供用户查看所提交的知识。

[0053] (3) 本体生成

[0054] 对于用于提交的每一条知识,系统可以通过解析知识内容,通过Jena工具包进行本体知识建模。

[0055] 本体的建模原语中包括:类的继承关系、数据类型属性 (Datatype Property)、关系属性 (Object Property)。系统解析具体的知识属性值及其类型,将其和本体建模原语相对应,同时借用Jena工具包,以OWL为本体标准格式,完成知识的本体建模,最后以本体文件的形式存储。

[0056] 例如,某流程类知识包含信息如表6所示

[0057] 表6 某流程类知识的信息

[0058]

流程名称	NAME
领域专业	专业M
领域用途	用途U
构建人	构建人X
核心参数	参数P
属性名A	属性值V

[0059] 其具体信息和OWL本体中建模原语的对应关系如图6所示:

[0060] 根据此对应关系,可以通过Jena工具包对知识进行详细的本体建模,以OWL格式文件保存。

[0061] 本发明说明书中未作详细描述的内容属于本领域专业技术人员公知的现有技术。

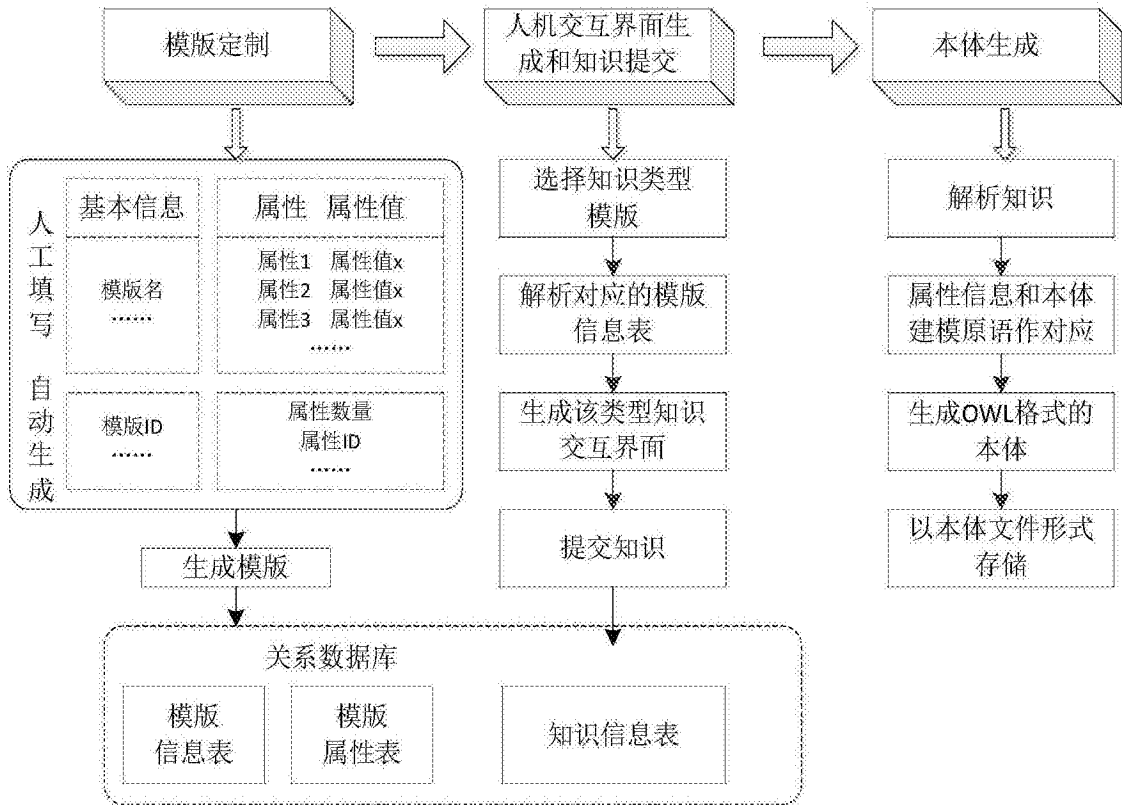


图1

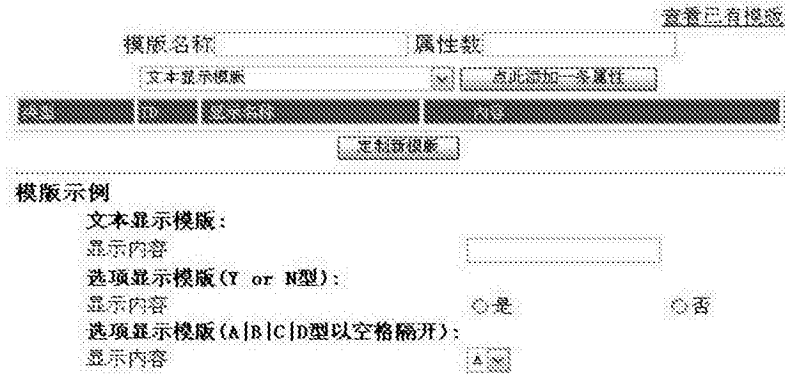


图2

查看已有模版

模版名称: 属性数:

选项显示模版 (a|b|c|i|j 型以空格隔开) 请从添加一条属性

序号	id	属性名	属性值
1	1	name	
2	2	sex	
3	3	age	10 20 25 30 35 40 >40

添加新模版

图3

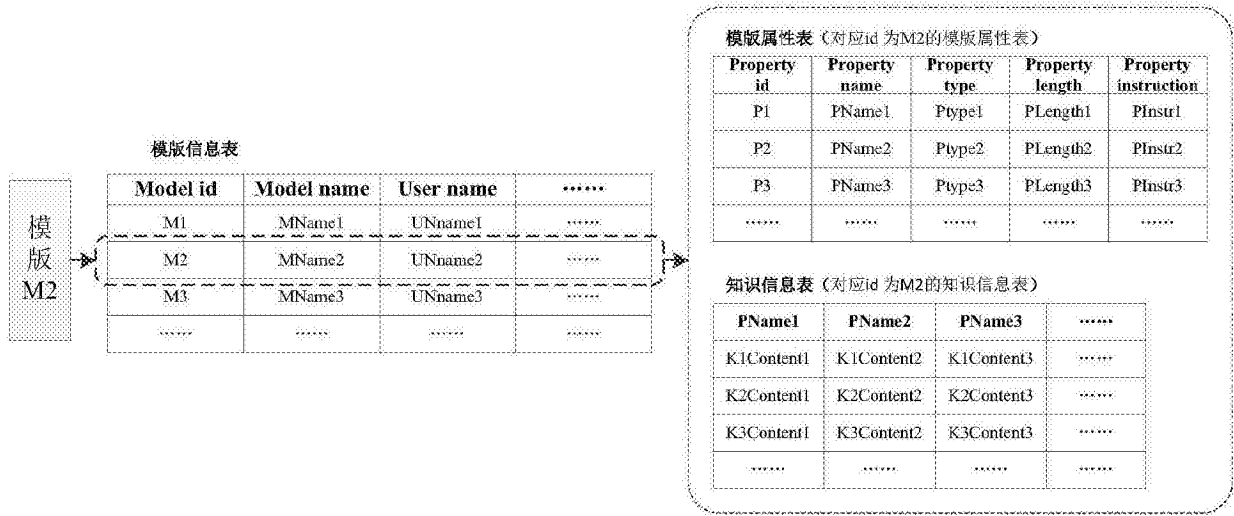


图4

领域模版: aaaa

提交专家: sys

名称	内容
name	<input type="text"/>
sex	<input type="radio"/> 男 <input type="radio"/> 女
age	<input type="text" value="10 20"/>

图5

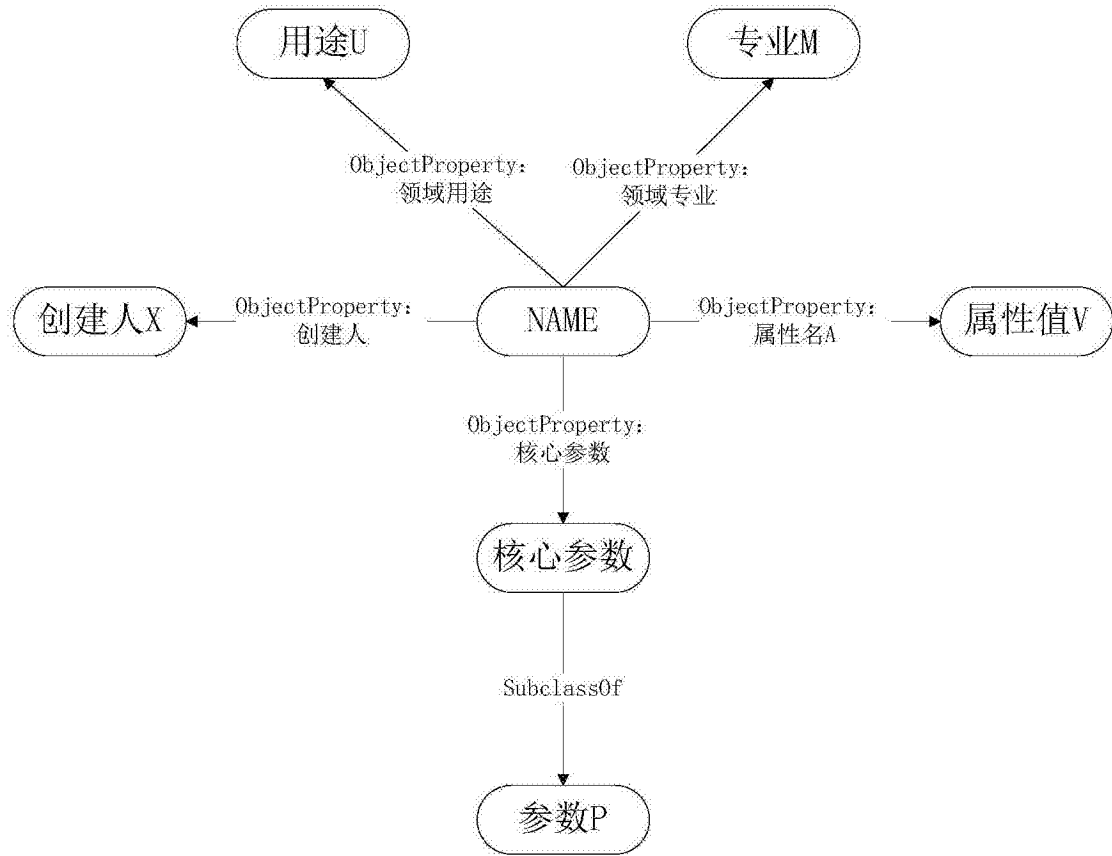


图6