## 知网知识库描述语言

1. **知网知识库描述语言2002版**
2. **总规定**
3. 任何一个概念的描述都以DEF= 为开始。任何一个概念中出现的所有义元或符号必须是在知网的Taxonomy中定义的义元或符号或者由知网知识库描述语言所规定的特定标识符。
4. 概念描述中的第一个义元必须指出该概念的最基本的意义，并用事件、实体、属性和属性值这四类义元中的一个标注出来。
5. 对于简单概念直接标注该概念的意义。
6. 利用动态角色与特征来标注复杂概念。
7. 属性类概念必须标明它的宿主。整体部分类型的概念必须标明该部分的整体。
8. 概念描述中定义的特性至少是一个，但也可以是多个，数量没有限制，只要内容是合理的且形式是合乎规范的就可以了。
   1. **KDML中的特定标识符。**

在知网知识库描述语言中允许使用以下7种标识符，它们是{ } : , = ; “，它们都是英文字符，它们的名称依次是：左括号、右括号、冒号、逗号、等号、分号、引号，其中左括号和右括号合称为大括号。具体见下表。

|  |  |  |
| --- | --- | --- |
| **符号** | **名称** | **功能简述** |
| { | 左括号 | 表示对一个概念描述的开始。 |
| } | 右括号 | 表示对一个概念描述的结束。 |
| : | 冒号 | 冒号后面的内容是对冒号前面义元的具体描述。 |
| , | 逗号 | 表示一个关系描述的结束。 |
| = | 等号 | 表示一个动态角色或特征所具有的具体的值。 |
| ; | 分号 | 分号表示某一概念是由若干个概念组合而成的组合型复杂概念。每个分号分割的部分必须是一个独立的完整的概念描述。 |
| “ | 引号 | 引号中的内容都是一些具有特殊意义的义元。 |

表2：KDML中的特定标识符

* 1. **几种特殊的指示符号**

1. 指示符号~

利用~进行描述的模式是:

{义元1:{义元2:动态角色或特征={~}}} －模式1

这种描述方式表示的是，义元1与义元2有关，义元1为义元2的一个具体动态角色的值。其中的~用来代替前面的义元1。通常情况下，义元1为实体类义元，义元2为事件类义元。

1. 指示符号?

利用?用来描述的模式是：

DEF={义元1:动态角色={?}} —模式2

这种描述方式表示在某一语义环境中，?所充当的动态角色的演员是一定会出现的，但是在这个孤立的概念中它并没有被体现出来。其中义元1一定是事件类义元。

1. 指示符号$

利用义元$进行描述的模式是：

DEF={义元1:动态角色={$}} —模式3

其中$用来充当某一个动态角色的演员。其中动态角色={$}表示这个概念所描述的对象是什么。义元1一定是事件类义元。

* 1. **简单概念的描述方法。**

直接标注该概念的意义。通常情况下的简单概念是指一个明确的事件，实体，属性或属性值，在概念中不包含任何的其它成分。

* 1. **复杂概念的描述方法。**

利用动态角色与特征来标注复杂概念。所谓的复杂概念是以事件为中心，除了事件中心本身以外还有一个或一个以上的动态角色，例如:

|  |  |
| --- | --- |
| 严禁 | 包含动态角色——方式 (manner) |
| 贷款 | 包含动态角色——所有物 (possession) |
| 盗墓 | 包含动态角色——来源 (source) |
| 复原 | 包含动态角色——原状态 (StateIni) |
| 呼救 | 包含动态角色——目的 (purpose) |

表3:事件的动态角色

在表示上述动态角色时它的书写格式是：

动态角色名称={某一概念描述}

* 1. **举例**

例1：“洗衣机”（实体类概念）

*自然语言描述：*

洗涤衣物的机器。

*KDML描述：*

DEF={tool|用具:{wash|洗涤:instrument={~},patient={clothing|衣物}}}

洗衣机是一个复杂的概念，在这里tool|用具、wash|洗涤、clothing|衣物等为义元，instrument、patient等均为关系，~为指示符号，{ } : = , 为KDML中的特定标识符。

它的定义以DEF=为开始，其中的第一个义元tool|用具为实体类义元，表示洗衣机是一个实体类的概念，它相当于模式1中的义元1。wash|洗涤是一个事件类的概念，它相当于模式1中的义元2。为了说明该实体与该事件之间的关系，运用~来代替义元1，从而说明义元2的工具（instrument）是义元1。另外，通过patient这个关系指出wash|洗涤这个事件的受事是clothing|衣物。

这段描述所表达的意义是直观的，即“洗衣机”是一种用具，是洗涤的用具，洗涤的受事是衣物。通过和它的自然语言描述相比较，不难看出，它已经进一步接近了自然语言。

在知网知识库描述语言2002版的概念描述中，所有的关系都是显性的，这种描述是没有歧义的。而不是象在知网知识库描述语言2000版中那样，使用符号来表示各种关系，容易引起歧义。这种显性的标注无论对标注、对理解还是对计算机计算而言，都能体现它的优越性。

例2：“出狱”（事件类概念）

*自然语言描述：*

走出监狱或不再受监禁。

*KDML描述：*

DEF={undergo|经受:content={release|释放:source={InstitutePlace|场所:domain={police|警},{detain|扣住:location={~},patient={human|人:modifier={guilty|有罪}}},{punish|处罚:location={~},patient={human|人:modifier={guilty|有罪}}}}}}

出狱是一个很复杂的事件类概念。在这个概念的描述中嵌套了“监狱”这个概念的完整定义，即{InstitutePlace|场所:domain={police|警},{detain|扣住:location={~},patient={human|人:modifier={guilty|有罪}}},{punish|处罚:location={~},patient={human|人:modifier={guilty|有罪}}}}。而在“监狱”的概念中又嵌套了“罪犯”这个概念的完整定义，即{human|人:modifier={guilty|有罪}}。

在知网知识库描述语言2002版中通过特定标识符的运用，使得对概念的描述呈现了一种多层嵌套的立体化的格局，从而增强了这种描述语言对概念描述的能力。不同的层次的之间又通过不同的关系指出它们之间的关系，这样便于意义的计算。同样的概念如果使用知网知识库描述语言2000版根本是无法表达的，即使能够通过一些义原和符号将这样的概念描述出来，在概念的程序解析上仍然存在能以处理的问题。

可见，知网知识库描述语言2002版通过显性表示各类义原之间的关系，采用立体化的多层嵌套的格局，极大地增强了这种描述语言的表达能力和计算能力。

1. ****应用KDML进行计算****

知网认为知识是一个系统，是一个包含着各种概念与概念之间的关系，以及概念的属性与属性之间的关系的系统。因此，知网对于概念的描述着力体现的是概念与概念、概念的属性与属性之间的相互关系。知网知识库描述语言是一种面向计算机的、可以进行计算的描述语言，通过计算我们应该可以得到概念与概念，概念的属性与属性之间的关系。下面来看一些例子。

例1：

NO.=123127

W\_C=钟表

G\_C=N

E\_C=

W\_E=clocks and watches

G\_E=N

E\_E=

DEF={tool|用具:{tell|告诉:content={time|时间},instrument={~}}}

例2：

NO.=123130

W\_C=钟表匠

G\_C=N

E\_C=

W\_E=watch maker

G\_E=N

E\_E=

DEF={human|人:HostOf={Occupation|职位},{repair|修理:agent={~},patient={tool|用具:{tell|告诉:content={time|时间},instrument={~}}}}}

在上面的例子中，例1表示：钟表是一种用具，是告诉时间的工具。例2表示：钟表匠一个人，是一个有职位的人，他的工作是修理例1所描述的用具。从中我们不难看出，知网知识库描述语言在概念表达上具有直观、可读性较好的特点，在描述能力上已经具备了描述复杂概念的能力。而最重要的是它非常方便于意义的计算。而意义的计算是面向计算机的语义资源的最关键的任务和衡量其质量的最主要的标准。尽管知网知识库描述语言对概念的描述是孤立的、静止的，但是我们可以通过运算来激活这些概念之间的动态关系。

从这个例子中，可以看到我们的做法。尽管对每一个概念的描述中没有描述它与其它概念之间有什么样的关系，但是例2的概念中嵌套了例1的完整定义，这两个概念也因此而产生了特定的关系。

例3：

NO.=007509

W\_C=表针

G\_C=N

E\_C=

W\_E=hand

G\_E=N

E\_E=

DEF={part|部件:whole={tool|用具:{tell|告诉:content={time|时间},instrument={~}}}}

例4：

NO.=085366

W\_C=时针

G\_C=N

E\_C=

W\_E=hour hand

G\_E=N

E\_E=

DEF={part|部件:whole={tool|用具:{tell|告诉:content={time|时间:TimeSect={hour|时}},

instrument={~}}}}

例3和例4描述的都是部件－整体关系。我们知道，表针和时针都是钟表的一部分。在例3中，例1所描述的概念的整个DEF全部嵌套在例3中。而例4中的整体则是例1所描述的概念的DEF的变形。也就是说，概念与概念之间的关系并不是通过简单的字符串匹配就可以得到的，而是需要将DEF解析并通过一定的法则而最终确定的。

目前，有两个利用知网进行这种意义的计算的应用，一是词语的相似度的计算，一是词语的相关性的计算。前者是刘群研发的[3]，后者是董强研发的，称为《相关概念场》。这两个软件包的一个共同点是它们都是可以任意测试的，而不是仅仅只有几个“漂亮”的例子而已。而它们的不同之处在于，前者利用的是知网知识系统2000版，其中的描述语言是知网知识库描述语言2000版，而后者利用的是知网知识系统2002版。在前面我们说过，知网知识库描述语言2000版在描述能力上存在一定的不足之处，因此它不利于意义的计算，相信如果使用知网知识系统2002版进行相似度计算会得到更好的效果。

1. ****结束语****

知网知识库描述语言的诞生是由于知网知识库中概念描述的需要，知网的发展过程也正是知网知识库描述语言的不断成长、成熟的过程。目前经过对中英文两种语言各8万多概念的描述，证明它没有歧义，能够提高概念描述的可读性，增强了概念描述的能力，使其描述明确化、立体化，为概念的关系计算提供了更好的途径。世界总是处在不断的变化之中，知网知识库描述语言也不例外，它必然会随着知网知识体系的发展而完善，成为表达能力更强、计算能力更强、语法更严密的一种知识描述语言。

NO.=111003

W\_C=医生

G\_C=N

E\_C=

W\_E=doctor

G\_E=N

E\_E=

DEF={human|人:HostOf={Occupation|职位},domain={medical|医},{doctor|医治:agent={~}}}

NO.=111022

W\_C=医院

G\_C=N

E\_C=

W\_E=hospital

G\_E=N

E\_E=

DEF={InstitutePlace|场所:domain={medical|医},{doctor|医治:content={disease|疾病},location={~}}}

例如：“车夫”

*自然语言描述：*

指以驱车或拉车为职业的人。

*KDML描述（2000版）：*

DEF=human|人,#occupation|职位,\*drive|驾驭,#LandVehicle|车

*说明：*

这个标注中义元drive|驾驭和义元LandVehicle|车之间顺序是绝对不可以调换的，它们的这种顺序说明车是驾驭的受事。

例如：“北麓”

*自然语言描述：*

北面的山脚。

*KDML描述（2000版）：*

DEF=part|部件,%land|陆地,base|根

*说明：*

其中有一个很明显的意义“北”没有被描述出来，这是因为，“北”这个概念在这里修饰的是山脚，如果按照语法规则，只能将north|北这个义元放在最后，但是这时这个义元究竟是修饰前面三个义元中的哪一个义元就很难把握了。

对于知网知识库描述语言2000版就介绍这么多。下面对知网知识库描述语言2002版进行介绍。

NO.=127567

W\_C=罪犯

G\_C=N

E\_C=

W\_E=offender

G\_E=N

E\_E=

DEF={human|人:modifier={guilty|有罪}}

NO.=047772

W\_C=监狱

G\_C=N

E\_C=

W\_E=prison

G\_E=N

E\_E=

DEF={InstitutePlace|场所:domain={police|警},{detain|扣住:location={~},patient={human|人:modifier={guilty|有罪}}},{punish|处罚:location={~},patient={human|人:modifier={guilty|有罪}}}}