

基于 HowNet 的航空术语语义知识库的构建*

张桂平, 刁丽娜, 王裴岩

(沈阳航空航天大学 知识工程研究中心, 辽宁 沈阳 110136)

摘要: 语义知识库的构建是自然语言处理基础性工作, 对于语言信息的处理有重要的作用, 但面向特定领域的语义知识库的构建还是一个难点。该文在分析了航空术语的基本特点的基础上, 根据 HowNet 和 KDML 描述语言构建了面向航空领域的术语语义知识库, 并在构建航空术语知识库的过程中总结形成了构建航空术语知识库的基础规则、动态角色/特征的选择规则。在文章最后对所构建的术语进行了相似度的计算, 取得了较好的结果。

关键词: 航空术语; HowNet; 语义知识库; KDML

中图分类号: TP391

文献标识码: A

Construction of Aviation Terminology Semantic Knowledge Base Based on HowNet

ZHANG Guiping, DIAO Lina, WANG Peiyan

(Knowledge Engineering Research Center Shenyang Aerospace University, Liaoning Shenyang, 110136, China)

Abstract: Semantic knowledge base construction is a basic work for the natural language processing. It's a great of help for language information processing. But it's difficult to construct knowledge base for specific domain. In this paper, based on the analysis of the basic characteristics of the aviation terminology, we built an aviation terminology semantic knowledge base according to HowNet and KDML, and summarize the foundation rules, Event Role and Features selection rules for building aviation terminology knowledge base during the process of building aviation terminology knowledge base. At the end of the article, we calculate the similarity between the built terminology, and achieves a good result.

Key words: Aviation Terminology; HowNet; Semantic Knowledge Base; KDML

1 引言

语义知识库(如机器词典、句法规则库等)是自然语言处理系统的重要组成部分,其规模与质量是自然语言处理系统成败的关键^[1]。因此构建语义知识库成为自然语言处理的基础工作。

目前国内外成功构建了许多的大规模语义知识库包括 WordNet^[2]、MindNet^[3]、FrameNet^[4]、Cyc 常识知识库^[5]、YAGO^[6]等,但大多都以自然语言形式进行表述,在自然语言处理系统中不易对其进行处理和计算。

而 HowNet^[7]利用义原揭示了概念之间的关系,利用 KDML 描述语言^[8]将概念的自然语言的表述形式描述成计算机可读的语言,这种描述形式使得计算机处理系统更容易进行相似度计算^[9]、相关度计算^[10]、倾向性计算等^[11],因此,HowNet 被广泛的应用于自然语言处理的各个领域。

但是现有的大规模知识库包括 HowNet 大都关于通用领域知识,不能满足特定领域的自然语言处理系统对问题求解的知识需求。为了更好的处理面向特定领域的信息,需要获取专业领域知识,即构建领域知识库。

本文针对航空领域在 HowNet 基础上进行了扩

* **收稿日期:** **定稿日期:**

基金项目: 国家“十二五”科技支撑计划项目(No. 2012BAH14F00)

作者简介: 张桂平(1962—),女,教授,主要研究方向为自然语言处理、机器翻译;刁丽娜(1988—),女,硕士,主要研究方向为自然语言处理;王裴岩(1983—),男,博士,主要研究方向为自然语言处理与人工智能。

展, 构建了航空领域的术语语义知识库, 包括 1000 条术语定义描述信息, 并总结了航空术语构建的基础规则和动态角色/特征的选择规则。

本文第二节简要介绍了 HowNet 和 KDML 描述语言, 第三节详细介绍了航空术语语义知识库的术语构建步骤, 对构建过程中使用的基础规则和动态角色/特征选取规则进行了详细介绍, 第四节根据本文构建的航空术语语义知识库对所构建的航空术语进行了相似度计算, 最后是本文的结论。

2 HowNet 与 KDML 描述语言简介

“知网”从 1988 年提出至今已有 20 多年的历史, 被公认为规模最大, 收录词语最多, 提供信息最多的语义词典^[12]。KDML(Knowledge Database Mark-up Language)是由“HowNet”研发并使用的一种知识描述规范体系, “知网”的成功更加证实了 KDML 具有很强的描述能力。香港科技大学颜国伟博士给予了 KDML 很高的评价, 他指出: “知网的知识表达模式是针对计算机的信息处理特点而制定的”^[13]。

2.1 “知网”描述概念总规定

在“知网”2006 版中共描述了 84826 个中文概念, 对于概念的描述需要满足概括性与一般性的特点, 同时也要保证概念描述的一致性与准确性要求, 为此给出如下规定^[8]:

(1) 任一概念的描述都以“DEF=”为开始。任一概念中出现的所有义原或符号必须是在知网的 Taxonomy 中定义的义原或符号或者由知网知识系统描述语言所规定的特定标识符;

(2) 概念描述中的第一个义原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原中的一个标注出来;

(3) 对于简单概念直接标注该概念的意义;

(4) 利用动态角色与特征来标注复杂概念;

(5) 属性类概念必须标明它的宿主;

(6) 整体部分类型的概念必须标明该部分的整体;

(7) 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。

2.2 KDML 描述语言特定标识符

KDML 允许使用以下 7 种标识符, 它们都是英文字符, 包括“{”、“}”、“:”、“=”、“;”、“,”和“””, 其中“{”表示一个概念描述的开始, “}”表示一个概念描述的结束, “:”后边的内容是对前边义原的具体描述, “,”表示一个关系描述的结束, “=”表示一个动态角色或特征所具有的具体的值, “;”表示某一概念是由多个概念组合而成的复合型复杂概念, 每个“;”分割的必须是一个独立、完整的概念, “””引号中的内容是具有特殊意义的义原^[8]。

2.3 KDML 描述语言特殊指示符号

除上述 7 种特定标识符之外, KDML 描述语言在对概念的描述还存在三个特殊的指示符号: “~”、“?”和“\$”^[8]。

利用“~”进行概念描述的特定形式是:

$$DEF = \{Y_1; \{Y_2; EF = \{\sim\}\}\}$$

其中, Y_1 表示义原, EF 表示动态角色与特征。

这种描述方式表明 Y_1 与 Y_2 有关, Y_1 为 Y_2 的动态角色的值, 用“~”代替前边的义原 Y_1 。

利用“?”进行描述的模式是:

$$DEF = \{Y_1; EF = \{?\}\}$$

这种描述表示在某一语境中“?”所充当的动态角色的演员一定会出现, 但在此孤立的描述中并没有体现出来。

利用义元“\$”进行描述的模式是:

$$DEF = \{Y_1; EF = \{\$\}\}$$

“\$”用来充当某一个动态角色的演员。“\$”代表这个概念所描述的对象。

2.4 动态角色与特征

KDML 通过 89 种动态角色与特征来描述概念所表述的义原间的关系, 下面列出其中使用频度较高的动态角色/特征的具体含义^[8]:

(1) “patient” (受事) 表示行动的事件类型中“变关系”、“变状态”、“变属性”和“使-动”四类事件中的充当“被改变”这一功能的实体。

(2) “content” (内容) 表示“泛动”、“精神状态”和“变精神”等类事件中被涉及的实体, 它与“受事”的区别在于: “内容”不承受“改变”; 也可表示实体的属性等。

(3) “agent” (施事) 表示行动的事件类型中“变关系”、“变状态”、“变属性”和“使-动”四类事件中的充当“变”这一功能的实体。

(4) “condition” (条件) 事件发生或进行的条件。

(5) “modifier” (描述) 为被修饰对象增加某种属性值。

(6) “CoEvent” (对应之事件) 说明一个实体类的概念与一个事件享有完全相同的义原。

(7) “whole” (整体) 表示“蕴涵关系”或表示行动事件中“选择”类事件中的实体的整体。

(8) “AccordingTo” (根据) 事件发生或进行的根据。

(9) “time” (时间) 事件发生的时间。

(10) “location” (处所) 事件发生的空间。

(11) “RelateTo” (相关) 表示一个概念与什么样的信息有关, 但是这种关系可能是不清晰的。

(12) “host” (宿主) 属性的主人即宿主。

(13) “purpose” (目的) 事件发生进行的目的。

(14) “instrument” (工具) 事件发生或进行所依赖的工具。

(15) “material” (材料) 事件发生或进行所依赖的材料。

(16) “whole” (整体) 表示“蕴涵关系”或表示行动事件中“选择”类事件中的实体的整体。

从上述对“知网”和 KDML 描述语言的介绍, 可以知道 KDML 描述语言的描述能力很强, 义原关系清楚, 而且具有计算机可读性、直观易懂、易于计算的优点, 是比较理想的知识描述语言。

3 语义知识库的构建

3.1 航空术语语义知识库的构建步骤

航空术语语义知识库相对于通用领域的语义知识库来说具有很强的专业领域特性, 对于此类知识库的构建现在主要有两种方法: (1) 直接构建该专业领域的知识库, (2) 利用现在比较成熟的知识库为基准进行该领域知识的扩建^[12]。本文的知识库构建选择上述第二种方法, 以“知网”为基准向航空领域知识进行扩建。

这主要基于以下两点考虑: 首先, 所需描述的航空术语是以自然语言的形式描述, 包含大量常识概念, 而“知网”涵盖了这些概念, 所以不需要再重新构建。

其次, “知网”揭示了概念内部属性之间的关系, 而本知识库构建的目的是实现航空领域的术语内部

的语义关系的知识库。这与知网类似, 都是以描述术语的语义特性为重点。

本文构建航空术语语义知识库所采用的术语来源于航空百科辞典, 其中包括术语名称和自然语言表述的术语定义, 形式如下:

“刹车装置”: “利用摩擦来制动机轮转动的装置”。

航空术语语义知识库的构建过程主要包括三部分: 术语预处理、义原获取、义原整合。

首先, 术语预处理的主要任务是将术语定义进行分词, 根据汉语的语言特点, 中心词大多在句尾, 并且考虑到“知网”所涵盖的词语量比较大, 所以依据“知网”词表对术语定义进行最大逆向匹配分词。

例如: 术语“刹车装置”航空词典中给出的定义是: “利用摩擦来制动机轮转动的装置”。分词结果是: “利用 摩擦 来 制动 机 轮 转动 的 装置”。

第二, 义原获取是根据分词以后的术语词语集中选择核心词语, 并从“知网”的义原库中提取与之相关的义原。上述例子从知网中获得的核心理语和相关义原是:

摩擦: “rub|摩擦”、制动: “TurnOff|止动”、转动: “rotate|转动”、机: “machine|机器”、装置: “tool|用具”。

第三, 义原整合是根据已有义原, 选择合适的动态角色/特征, 根据描述规则将义原进行组合, 形成完整的术语表述形式, 是航空术语库构建过程中的主要过程。可以得到上述例子的描述形式是:

```
DEF={tool|用具:{TurnOff|止动:ResultEvent={rotate|转动:agen  
t={machine|机器}},RelateT={rub|摩擦},instrument={~}}
```

从上述步骤可以看出要将义原按照一定的规则进行整合, 形成规范的描述形式。其中基础规则和动态角色/特征选取规则是十分重要的, 因此, 在 3.2 节中详细介绍了航空术语语义知识库构建的基础规则。

3.2 基础规则

在术语库的构建过程中根据航空术语的特点, 结合“知网”的七条总规定, 延伸出符合航空术语知识库构建的五条基础规则:

1、整体部分类型的概念必须标明该部分的整体, 如果明确属于整体的哪一部分, 要用“PartPosition={ }”表示出来。

例如：

安全结构：DEF={part|部件:modifier={important|重要},PartPosition={bone|骨},whole={aircraft|飞行器}}

2、如果术语的第一个义原是“fact|事情”和“affairs|事务”必须用动态角色“对应之事件 CoEvent={ }”来描述具体事件。

例如：

程序制导：DEF={fact|事情:CoEvent={guide|引导:instrument={software|软件:{control|控制:instrument={~},AccordingTo={software|软件:{MakeAppointment|约定:instrument={~},content={time|时间}}}},patient={weapon|武器:{shoot|发射:patient={~}}}}

3、术语的第一个义原“tool|用具”后一定要有事件类的义原和“instrument={~}”来表示出具体是什么用途的工具。

例如：

测试设备：DEF={tool|用具:{check|查:instrument={~}}}

4、描述术语实体所起作用、功效的时候用“BeAble|能够”来描述。事件义原的目的用动态角色“purpose”来描述。

例如：

鞭状天线：DEF={tool|用具:{BeAble|能够:agent={~},content={curved|弯},modifier={linear|线}{LengthLong|长}{fine|纤},{transmit|传送:instrument={~},patient={electricity|电}}}
校准设备：DEF={tool|用具:{check|查:purpose={amend|改正},instrument={~}}}

5、任意一个概念的描述，两个义原都不能直接连接，必须由动态角色描述出两个义原之间的关系，除非在利用“~”的情况下，形式如下：

DEF = {Y₁:{Y₂:EF={~}}}

例如：

自动测试设备：DEF={tool|用具:{check|查:instrument={~},modifier={automatic|自动}}}

本术语库的所有概念描述都是基于“知网”的七条总规则，和在此提到的五条基础规则下进行的。

在术语描述中如何选择合适的动态角色/特征是构建过程中的难点，在本术语库的构建过程中根据航空术语的特点对一些动态角色的用法进行了新的规定，这些规定可以分为五类，在 3.3-3.7 节详细

介绍，包括：动态角色的使用、易混淆动态角色的区分、动态角色的选择、特殊义原的用法和特殊情况的表述形式。具体规则如下：

3.3 动态角色使用

1、“belong”和“domain”都有表示归属领域的意思，但是“domain”有固定搭配的义原（这些义原属于“Secondary Feature”，只能接在“domain”后），而且所表示的范围都比较大，表示某一特定领域，“belong”表示成员归属一个整体，后边所接可以是完整的定义描述，核心义原（一条定义描述的第一个义原）只要求是实体。格式如下：

DEF={Y₁:belong={Y₂}}
s.t. Y₂ is {(entity|实体):}

DEF={Y₁:domain={Y₃}}
s.t. Y₃ is "domain|领域"

“s.t.”代表特定的情况和条件。

例如：

远距操纵炮塔系统：DEF={facilities|设施:{control|控制:instrument={~},patient={facilities|设施:domain={military|军},{firng|射击:instrument={~}}},manner={far|远}}}
金属中气体分析：DEF={fact|事情:CoEvent={analyze|分析:content={gas|气:location={location|位置:belong={metal|金属}}}}

2、“PatientAttribute”（受事属性）表示改变实体的某一属性。结构如下：

DEF₁={Y₁:PatientAttribute={Y₂}}
s.t. Y₁ is {(event|事件)}
Y₂ is {(Attribute|属性)}

例如：

压气机：DEF={machine|机器:{MakeHigher|提高:PatientAttribute={Strength|力量:host={gas|气},scope={press|按压}},instrument={~}}}

3、“ResultEvent”表示“使之动”等事件所导致的并由其受事施行的事件，多用于“事件类义原+事件类义原”的情况，即“动词+动词”。结构如下：

DEF={Y₁:ResultEvent={Y₂}}
s.t. Y₁ is {(event|事件)}
Y₂ is {(Attribute|属性)}

例如：

导航：DEF={guide|引导:ResultEvent={VehicleGo|驶}}

4、“CoEvent”（对应之事件）后只能接事件类义原。结构如下：

DEF₁={Y₁:CoEvent={Y₂}}
s.t. Y₂ is {(event|事件)}

例如：

卫星通信：DEF={fact|事情:CoEvent={communicate|交流},ResultTo={aircraft|飞行器}}

5、“modifier”（描述）为被修饰对象增加某种属性值。后接的义原一定是属性值类的。结构如下：

DEF={Y₁:modifier={Y₂}}
s.t. Y₂ is {(AttributeValue|属性值)}

例如：

高超声速风洞：DEF={facilities|设施:{check|查:content={wind|风},instrument={~}},{speed|速度:modifier={high|高:degree={very|很}},ContentCompare={sound|声}}

6、“manner”（方式）描述事件类义原，其值的核心义原是属性值类的义原。结构如下：

DEF={Y₁:manner={Y₂}}
s.t. Y₁ is {(event|事件)}
Y₂ is {(AttributeValue|属性值)}

例如：

攀升：DEF={BecomeMore|增多:manner={continuous|连续}}

7、“means”（手段）描述事件类义原，其值是事件类的义原。结构如下：

DEF={Y₁:means={Y₂}}
s.t. Y₁ is {(event|事件)}
Y₂ is {(event|事件):}

例如：

调配：DEF={produce|制造:means={mix|混合}}

8、“host”（宿主）表示属性的主人。“host”的值多为实体类义原，也有属性类的义原。结构如下：

DEF = {Y₁:host = {Y₂}}
s.t. Y₁ is {(Attribute|属性)}
Y₂ is {(Attribute|属性)}
or {(entity|实体)}

例如：

长度限制：DEF={Boundary|界限:host={Distance|距离}{Length|长度}{Width|宽度}}

9、“patient”（受事）表示行动的事件类型中“变关系”，“变状态”，“变属性”，“使之动”四类事件中的充当“被改变”这一功能的实体。

例如：

露头角：DEF={CauseToAppear|显现:patient={Ability|能力}}

上述例子中“Ability|能力”是属性类的义原，这种情况在本术语库的构建中不采用“patient=”而用“content=”，属性的宿主用 host={ }表示，“content={ Ability|能力: host={ } }”。结构如下：

DEF = {Y₁:patient={Y₂}}
s.t. Y₁ is {(event|事件)}
Y₂ is {(entity|实体)}

例如：

飞溅结构：DEF={part|部件:{obstruct|阻止:ResultEvent={jet|喷出:patient={water|水}},instrument={~}},PartPosition={bone|骨}, whole={aircraft|飞行器}}

3.4 易混淆动态角色的区分

1、“patient”、“content”、“PatientAttribute”的用法和区别：“patient”承受“改变”，“content”不承受“改变”；如果属性值改变则用“PatientAttribute”，“PatientAttribute”后之接“属性值”类的义原；如果属性值没有变化则用“content”来表示，“content”后接“实体”类和“事件”类义原。结构如下：

$DEF_1 = \{Y_1 : patient = \{Y_2\}\}$
 $DEF_2 = \{Y_3 : PatientAttribute = \{Y_4\}\}$
 $DEF_3 = \{Y_5 : content = \{Y_6\}\}$
 s.t. Y_1, Y_3, Y_5 is {(event|事件)}
 Y_2 is {(event|事件):}
 Y_4 is {(Attribute|属性):}
 Y_6 is {(Attribute|属性) or (event|事件):}

例如:

空气调节系统试飞: $DEF = \{fact|事情: CoEvent = \{experiment|实验: content = \{fly|飞\}, manner = \{try|尝试\}, purpose = \{check|查: patient = \{part|部件: \{adjust|调整: instrument = \{\sim\}, PatientAttribute = \{Temperature|温度\}\}\}, whole = \{aircraft|飞行器\}\}\}$

2、“target”（目标）事件中涉及到的，但没有被改变的实体，与“patient”的区别在于不承受改变，而且带有倾向性和目的性。

$DEF = \{Y_1 : target = \{Y_2\}\}$
 s.t. Y_1 is {(event|事件)}
 Y_2 can't be changed.

例如:

磁性材料: $DEF = \{physical|物质: HostOf = \{Performance|性能: cope = \{attract|吸引: target = \{metal|金属\}\}\}\}$

3、“partner”指两个或多个在一起，“SourceWhole”指一个大的整体去掉小部分对整体几乎无影响。

例如:

环山: $DEF = \{BeNear|靠近: partner = \{land|陆地: modifier = \{protuding|凸\}\}\}$
 退伍: $DEF = \{withdraw|退出: SourceWhole = \{army|军队\}, domain = \{military|军\}\}$

3.5 动态角色选择

1、“AlterPossession|变领属”的事件义原的被领属者用动态角色“possession”和“patient”表示，结构如下:

$DEF = \{Y_1 : \{Y_2 : patient/possession = \{Y_3\}\}\}$
 s.t. Y_2 is {(AlterPossession|变领属)}

两者区别在“possession”表示 Y_1 是 Y_3 的拥有者，而“patient”只表示 Y_3 是 Y_2 动作的受事，不一定是属于 Y_1 。

例如:

电力公司: $DEF = \{InstitutePlace|场所: domain = \{economy|经济\}, \{provide|供: agent = \{\sim\}, possession = \{electricity|电\}\}\}$
 信息员: $DEF = \{human|人: \{provide|供: agent = \{\sim\}, patient = \{information|信息\}\}\}$

2、“change|变”如果是属性变化用“StateIni”（初始状态）、“StateFin”（终状态），如果是实体变化则用“experiencer”（经验者）表示事件的主体，结构如下:

$DEF_1 = \{Y_1 : \{change|变: StateIni = \{DEF_2\}, StateFin = \{DEF_3\}\}\}$
 s.t. DEF_2, DEF_3 is {(SituationValue|状况值)}

例如:

遇难呈祥: $DEF = \{change|变: StateFin = \{fortunate|好运\}, StateIni = \{unfortunate|不幸\}\}$

$DEF = \{Y_1 : \{change|变: experiencer = \{Y_2\}\}\}$
 s.t. Y_2 is {(entity|实体)}

例如:

振荡: $DEF = \{change|变: experiencer = \{electricity|电\}, time = \{time|时间: modifier = \{regular|定期\}\}\}$

(3)“separate|分离”如果是分离出去的东西对原来的整体影响不大（例如：从湖里分离出一杯水）则用“SourceWhole”（来源整体），如果分离出去的东西对原来的整体数量影响比较大（例如：从一杯水中分出半杯）则用“patient”，如果是分离两个相对独立的整体（例如：分开两个人）则用“partner”（相伴体），结构如下:

$DEF = \{Y_1 : patient/SourceWhole/partner = \{Y_2\}\}$

例如:

光谱分析: $DEF = \{fact|事情: CoEvent = \{analyze|分析: content = \{phenomena|现象: \{forming|形成: PatientProduct = \{\sim\}\}, means = \{separate|分离: SourceWhole = \{lights|光\}\{sound|声\}, patient = \{Color|颜色\}\}\}\}\}$
 隔声结构: $DEF = \{part|部件: \{separate|分离: patient = \{sound|声\}\}, PartPosition = \{bone|骨\}, whole = \{aircraft|飞行器\}\}$
 分离速度: $DEF = \{Speed|速度: host = \{physical|物质: modifier = \{hang|悬挂\}\}, TimeSect = \{separate|分离: partner = \{implement|器具: \{hang|悬挂: instrument = \{\sim\}\}\}\}\}$

4、关于义原“feed|喂”所“喂”的东西用“patient”描述，接收所“喂”的东西的主体用“target”表示，

结构如下:

$$\text{DEF}=\{\text{feed|喂}:\text{patient}=\{Y_1\},$$

$$\text{target}=\{Y_2\}\}$$

$$\text{s.t. } Y_1, Y_2 \text{ is } \{(\text{entity|实体})\}$$

例如:

铡草: DEF={break|折断:patient={physical|物质:{feed|喂:patient={~},target={animal|兽}}}}

3.6 特殊义原的用法

1、“direction|方向”属于实体类义原，“direction|方向”在表示实体前进方向的时候具有属性类义原的特性，在这种情况下可以把“direction|方向”认为是属性类原后边用“host”标明其宿主。

例如:

导向: DEF={direction|方向:host={information|信息}{thinking|思想},scope={guide|引导}}

2、“able|能”：是属性值类的义原，表示一种能力，用于描述实体所具有的某种特性，该特性与事件有关不能直接用属性值来描述。结构如下:

$$\text{DEF}=\{Y_1:\text{modifier}=\{\text{able|能}:\text{scope}=\{Y_2\}\}\}$$

$$\text{s.t. } Y_1 \text{ is } \{(\text{entity|实体})\}$$

例如:

旋翼: DEF={part|部件:modifier={able|能:scope={rotate|转动}},PartPosition={wing|翅},whole={aircraft|飞行器}}

3、“own|有”表示属于某一实体的东西，不一定是实体也可能包括知识等。与“OfPart”的区别不一定是这个实体的一部分，结构如下:

$$\text{DEF}=\{Y_1:\{\text{own|有}:\text{possessor}=\{~\},$$

$$\text{possession}=\{Y_2\}\}\}$$

$$\text{s.t. } Y_1 \text{ is } \{(\text{entity|实体})\}$$

$$Y_2 \text{ belong } Y_1$$

$$\text{DEF}=\{Y_3:\text{OfPart}=\{Y_4\}\}$$

$$\text{s.t. } Y_4 \text{ is a part of } Y_3$$

例如:

园艺家: DEF={human|人:domain={agricultural|农},{own|有:possession={knowledge|知识:concerning={FlowerGrass|花草}{vegetable|蔬菜}},possessor={~}}}

鸭式飞机: DEF={aircraft|飞行器:{fly|飞:instrument={~}},OfPart={wing|翅:modifier={horizontal|横}{small|小}}}

4、“~”多用于描述实体与事件之间的关系的关系，即“做...的东西”，结构如下:

$$\text{DEF}=\{Y_1:\{Y_2:\text{EF}=\{~\}\}$$

$$\text{s.t. } Y_1 \text{ is } \{(\text{entity|实体})\}$$

$$Y_2 \text{ is } \{(\text{event|事件})\}$$

“~”代表Y1，表示Y1是Y2的一个具体动态角色的值。

例如:

天线: DEF={tool|用具:modifier={linear|线},{transmit|传送:instrument={~},patient={electricity|电}}}

3.7 特殊情况的表述形式

1、如果描述某一实体，要具体描述实体的属性，用动态角色“modifier=”加义原“{PropertyValue|特性值:}”用动态角色“scope={}”描述具体属性。结构如下:

$$\text{DEF}=\{Y_1:\text{modifier}=\{\text{PropertyValue|特性值}:\text{scope}=\{Y_2\}\}\}$$

$$\text{s.t. } Y_1 \text{ is } \{(\text{entity|实体})\}$$

$$Y_2 \text{ is } \{(\text{Attribute|属性})\}$$

例如:

高密度物体: DEF={physical|物质:modifier={PropertyValue|特性值:scope={Density|密度},modifier={GreaterThanNormal|高于正常}}}

2、动词作为修饰成分来修饰名词，例如：“约定的位置”、“反射的光”和被动的结构，例如：“可以被检验的结构”。在本术语库中规定这类情况，如果要描述的实体有相应的动态角色如：“location|位置”和“location”、“material|材料”和“material”等可用如下结构来表示:

$$\text{DEF}=\{Y_1:\{Y_2:\text{相应动态角色}=\{~\}\}\}$$

例如:

续表

飞向约定的位置: DEF={ fly|飞:target={location|位置:{fixed|已定:location={~}}}}

如果没有相应的动态角色, 则用“RelateTo”描述。

DEF={Y₁:{Y₂:RelateTo={~}}}

例如:

利用反射的激光: DEF={use|利用:patient={lights|光:RelateTo={respond|回应}}}

3、如果定义的第一个义原是属性类的义原则必须用“host”指出其宿主, 如果其他的地方出现属性类义原, 如果能确定其宿主则用“host”指出, 如果不确定则不用标明。

4、涉及到很多航空、物理、化学、机械领域的专业知识, 而且许多涉及人名的专业术语, 例如: “多普勒导航雷达天线”, 很难准确描述。在本术语库的构建中对与专业知识和带人名的专业术语只需指出属于某一领域或与某一事物相关即可, 不需要深入描述专业的知识。

例如:

多普勒导航雷达天线: DEF={tool|用具:modifier={linear|线}, domain={physics|物理},{transmit|传送:instrument={~},patient={electricity|电}}}

4 术语相似度计算

根据上述构建规则构建了航空术语语义知识库, 从中选取了八组有代表性的术语描述形式见表 1:

表 1: 术语描述形式

术语	术语定义描述
第一组	
参数	DEF={symbol 符号:scope={Quantity 数量}}
几何参数	DEF={symbol 符号:scope={Quantity 数量},RelateTo={math 数学}}
参数测量	DEF={measure 测量:content={symbol 符号:scope={Quantity 数量}},patient={physical 物质}}
第二组	
测量技术	DEF={knowledge 知识:RelateTo={measure 测量}}
测试技术	DEF={knowledge 知识:RelateTo={check 查}}
探测技术	DEF={knowledge 知识:RelateTo={LookFor 寻}}

测量装置	DEF={tool 用具:{measure 测量:instrument={~}}}
测量	DEF={measure 测量}
第三组	
温度误差	DEF={Rate 比率:host={entity 实体},RelateTo={Temperature 温度}}
误差	DEF={Rate 比率:host={entity 实体}}
温度测量	DEF={measure 测量:content={Temperature 温度:host={aircraft 飞行器}}}
第四组	
加工	DEF={produce 制造}
铸造	DEF={produce 制造}
半固态金属铸造	DEF={produce 制造:material={metal 金属:{produce 制造:PatientProduct={~},material={metal 金属:modifier={StateLiquid 液态}},modifier={StateSolid 固态}{StateLiquid 液态}},PatientProduct={artifact 人工物}}
切削加工专家系统	DEF={software 软件:{handle 处理:patient={problem 问题:RelateTo={cut 切削:domain={industrial 工业}}{produce 制造:domain={industrial 工业}},modifier={all 全}},modifier={Wisdom 智慧}}
第五组	
传感器	DEF={part 部件:whole={implement 器具},{sense 感觉:instrument={~}}}
导引头	DEF={part 部件:whole={implement 器具},{sense 感觉:instrument={~}}}
探测器	DEF={tool 用具:{LookFor 寻:instrument={~}}}
检测器	DEF={tool 用具:{measure 测量:instrument={~}}}
第六组	
陀螺	DEF={tool 用具:{measure 测量:instrument={~},content={Speed 速度:host={physical 物质}},RelateTo={Angle 角度}}}
激光陀螺	DEF={tool 用具:{measure 测量:instrument={~},content={Speed 速度:host={physical 物质}},RelateTo={Angle 角度}},RelateTo={lights 光}}
陀螺舵	DEF={part 部件:whole={weapon 武器:{shoot 发射:patient={~}},{drive 驾驭:instrument={~}},RelateTo={tool 用具:{recreation 娱乐:instrument={~}}}}
第七组	
飞机	DEF={aircraft 飞行器:{fly 飞:instrument={~}}}
隐身飞机	DEF={aircraft 飞行器:{fly 飞:instrument={~},manner={covert 隐秘}}}
军用飞机	DEF={aircraft 飞行器:{fly 飞:instrument={~}},domain={military 军}}
飞机电气系统	DEF={part 部件:{provide 供:possession={electricity 电}},PartPosition={bone 骨},whole={aircraft 飞行器}}

续表

第八组	
通信	DEF={fact 事情:CoEvent={communicate 交流}}
卫星通信	DEF={fact 事情:CoEvent={communicate 交流:instrument={aircraft 飞行器}}
通信设备	DEF={tool 用具:{communicate 交流:instrument={~}}}

对上述八组术语描述形式,利用参考文献 14 提出的相似度的计算方法^[14]进行相似度计算,相似度计算结果见表 2:

表 2: 相似度计算结果

第一组				
	几何参数	参数测量		
参数	0.9150	0.0000		
第二组				
	测试技术	探测技术	测量装置	测量
测量技术	0.9778	0.9778	0.2233	0.0000
第三组				
	误差		温度测量	
温度误差	0.9150		0.0000	
第四组				
	半固态金属铸造	铸造	切削加工专家系统	
加工	0.7040	1.0000	0.0000	
第五组				
	导引头	探测器	检测器	
传感器	0.9716	0.1844	0.1734	
第六组				
	激光陀螺		陀螺舵	
陀螺	0.9575		0.1478	
第七组				
	隐身飞机	军用飞机	飞机电气系统	
飞机	0.9150	0.9000	0.1280	
第八组				
	卫星通信		通信设备	
通信	0.9150		0.2436	

由上述相似度计算结果可以看出所构建的术语,使计算机基本能够理解人类对于该术语的解释,这也验证了本术语库构建的准确性。

5 结论

本文提出了一种基于“HowNet”的领域知识库构建方法,详细介绍了航空术语知识库的构建流程、

构建规则、和动态角色选取的规则,对“HowNet”进行了航空领域的扩展,使其在航空领域可以发挥更大的作用,从相似度的计算结果可以看出本术语库的构建方法取得了较好的结果。

将航空术语表示为计算机能够理解和处理的形式,使其能够作为判断航空领域的术语语义相似相关程度的依据,从而为面向航空领域的自然语言处理工作,以及航空术语间语义计算等工作提供了一个知识库资源。

由于本术语语义知识库是人工构建的,耗时耗力,效率较低,不能快速的构建大规模的知识库,现在计算机科学领域发展迅速,大数据的时代已经到来,对于知识库的需求会更大,因此下一步工作要解决高效地收集、存储和归纳处理知识,同时智能地分析知识,迅速、准确构建面向专业领域的语义知识库的问题。

参考文献:

- [1] 刘云,俞士汶,朱学锋,等.现代汉语虚词知识库的建设[J].语言文字应用,2005(1):130-136.
- [2] Fellbaum C. WordNet: An electronic lexical database[M]. MIT Press,1998.
- [3] Richardson S D, Dolan W B, Vanderwende L. MindNet: acquiring and structuring semantic information from text[C]//Proceedings of the 17th international conference on Computational linguistics Volume 2. Association for Computational Linguistics, 1998: 1098-1102.
- [4] Baker C F, Fillmore C J, Lowe J B. The berkeley framenet project[C]//Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics Volume 1. Association for Computational Linguistics, 1998: 86-90.
- [5] <http://www.opencyc.org/>,[DB/OL].
- [6] Suchanek F M, Kasneci G, Weikum G. Yago: a core of semantic knowledge[C]//Proceedings of the 16th international conference on World Wide Web. ACM, 2007: 697-706.
- [7] Dong Z, Dong Q. HowNet and the Computation of Meaning[M]. Singapore: World Scientific, 2006.

- [8] 董振东, 董强. KDML—知网知识系统描述语言[DB/OL]. http://www.keenage.com/html/e_index.html.
- [9] 刘群, 李素建. 基于《知网》的词汇语义相似度计算[J]. 中文计算语言学, 2002, 7(2): 59-76.
- [10] 李素建. 基于语义计算的语句相关度研究[J]. 计算机工程与应用, 2002, 38(7): 75-76.
- [11] 熊德兰, 程菊明, 田胜利. 基于 HowNet 的句子褒贬倾向性研究[J]. 计算机工程与应用, 2009, 44(22): 143-145.
- [12] 王莹莹, 白宇, 丁长林, 等. 面向语义检索的中医理论知识库构建方法的研究[J]. 中文信息学报, 2012, 26(5): 72-78.
- [13] 董振东, 董强. 面向信息处理的词汇语义研究中的若干问题[J]. 语言文字应用, 2001, 3: 27-32.
- [14] 夏天. 汉语词语语义相似度计算研究[J]. 计算机工程, 2007, 33(06): 191-194.