

DOI: 10.13382/j.jemi.2017.04.023

# 基于本体及 Web 文本的数控机床知识获取\*

刘锴锋 王红军 左云波

(北京信息科技大学 机电工程学院 北京 100192)

**摘要:**为了有效解决数控机床领域,由于知识没有结构化描述,导致知识获取效率低的问题以及实现知识重用和知识共享,通过对该领域 Web 文本中机床知识进行研究,分析文本结构特点,提出一种基于本体的数控机床知识抽取方法。对爬虫程序获取的文档进行预处理,通过模式匹配的方式抽取 Web 文本中存在上下位关系的语句,经过中文分词系统 ICTCLAS 分词处理后抽取概念,构建概念集合和概念树,最终构建领域本体并以 OWL 语言储存。实验中对随机选取的网页进行知识抽取,并采用对比实验,证明该方法能有效地对数控机床领域中半结构化和结构化 Web 文本信息进行获取。

**关键词:** 数控机床;知识获取;本体;中文分词

**中图分类号:** TP391      **文献标识码:** A      **国家标准学科分类代码:** 520.2070

## Knowledge acquisition of CNC machine tools based on ontology and web text

Liu Kaifeng Wang Hongjun ZuoYunbo

(Mechanical & Electrical Institute, Beijing Information Science & Technology University, Beijing100192, China)

**Abstract:** Aiming at Knowledge reuse and knowledge sharing and solving the problem of inefficiency of knowledge acquisition caused by the knowledge are not structured description, a method of knowledge acquisition based on ontology for CNC machine tools is proposed by studying the Web text in the field of machine tool knowledge and analyzing of text structure features. The original information of the Web text is obtained through the crawler program, and the sentences containing the hyponymy is extracted from the processed Web text information. The Institute of Computing Technology, Chinese Lexical Analysis System (ICTCLAS) is used to carry on the Chinese word segmentation. The concepts can be acquired after segmenting Chinese texts and then the concept set and the concept tree can be generated. At last, the ontology is generated and stored as OWL. In the experiment, the knowledge is extracted from the Web pages which is selected randomly, and the comparison experiment is carried out to prove that the method can obtain the semi-structured and structured Web text information in the field of CNC machine tools effectively.

**Keywords:** CNC machine tools; knowledge acquisition; ontology; Chinese word segmentation

## 1 引言

随着计算机技术和网络技术的不断发展和应用,大量数控机床的基本信息和检测维修案例及数据以 Web 文本的形式保存在网络中。然而由于网络信息组织格式主要为非结构化和半结构化,缺乏统一的规范,这些信息不能准确、快速的应用于数控机床检测和维修中。传统

处理 Web 信息的做法是人工阅读,搜索关键字等总结出相关信息,自动化水平低,耗费巨大的人力和物力,很大程度上延缓了知识获取速度。此外,传统方法中获取到的知识,重用性和共享性低,制约了知识的使用效率。

信息抽取是利用知识进行推理求解问题的前提,Grishman 作为 Proteus 工程的创建者,在 1997 年对信息抽取的概念进行了定义:“信息抽取涉及到为从文本中选择出的信息创建一个结构化的表示形式(比如数据

收稿日期:2017-01      Received Date: 2017-01

\* 基金项目:国家自然科学基金(51575055)、北京市教委科研计划(KM201611232020)、北京市重点实验室开放课题(KF20161123203)、北京市重点实验室开放课题(KF20161123201)项目资助

库)”<sup>[1]</sup>。传统信息抽取的方法有基于自然语言的方法,基于 HTML 结构提取,基于包装器归纳方式<sup>[2]</sup>。基于本体的信息抽取方法作为一种新的方式,越来越多的受到领域专家的关注。数控机床相关 Web 文本中包含大量专业词汇以及复杂的结构关系,采用基于本体的信息抽取方法,不仅能够清晰的表示知识对象之间的复杂性和相关性,还能为后期知识推理提供逻辑推理基础。因此,文本提出一种基于本体的数控机床知识获取和表示方法。

## 2 领域本体与 Web 文本

本体的概念最初起源于哲学领域。它在哲学中的定义为“对世界上客观存在物的系统的描述,即存在论”,是客观存在的一个系统的解释或说明,关心的是客观现实的抽象本质<sup>[3]</sup>。人工智能、计算机语言以及数据库理论学科将 Ontology 的概念从哲学领域中借用过来,并赋予了一些新的含义。美国 Stanford 大学的知识系统实验室的学者 Gruber 给出了 Ontology 一个较为广泛接受的定义,即 Ontology 是对概念化的明确描述,用于描述事物的本质<sup>[4]</sup>。本文研究本体的目的是为了获取数控机床领域知识,捕获领域内共同认可的术语,并且确定这些公认术语在不同层次的形式化模式上的相互关系,简单来说就是将知识转化为概念以及概念之间的关系。在基于本体的 Web 注释(OWA)形式化定义中,本体表示为  $O = \{C, A^c, R, A^r, H, X\}$ 。其中,  $C$  表示概念 (Classes) 集合;  $A^c$  表示本体概念中每个概念的属性集;  $R$  表示概念间关系 (Relations) 的集合;  $A^r$  表示关系集中每个关系的属性集;  $H$  表示概念层次;  $X$  表示公理集<sup>[5]</sup>。其中,概念之间有 4 种基本的关系  $P(x, y), K(x, y), A(x, y), I(x, y)$ <sup>[6]</sup>。

- 1)  $P(x, y)$  表示概念  $y$  是概念  $x$  的一部分,即整体与部分的关系 (part-of)。
- 2)  $K(x, y)$  表示概念  $y$  是概念  $x$  的子概念,即继承关系 (kind-of)。
- 3)  $A(x, y)$  表示概念  $y$  是概念  $x$  的一个属性,即属性关系 (attribute-of)。
- 4)  $I(x, y)$  表示概念  $y$  是概念  $x$  的一个实例,即实例关系 (instance-of)。

目前主流的知识表示方法有框架表示法,产生式表示法,谓词逻辑表示法,人工神经网络表示法等单一的知识表示方法<sup>[7]</sup>,为了强化知识表示能力,也有人将几种知识表示方法组合在一起使用,如基于对象混合型知识表示方法<sup>[8]</sup>。相较于这些方法,基于本体的表示方法,能够结构化表示领域知识,表达能力强,知识表达清晰,逻辑严谨,利于扩充和维护,此外,由于数控机床设计的知识范围广泛,语义关系复杂,基于本体的表示方法能够满足知

识传递和知识共享的准确性要求。

描述本体的语言从形式上可以分为两种,一种是传统描述语言,另一种是基于 Web 的描述语言。OWL 语言是一种 W3C 组织推荐的以 DAML + OIL 为基础,基于 Web 的描述语言标准,使用与 RDF 相同的语法来表示本体并且允许用机器可以理解 and 表达的方式来进行语义的描述。包括 OWL-lite、OWL-DL、OWL-full 三种子语言<sup>[9]</sup>。

与数控机床相关的 Web 文本中包含大量的半结构化知识,其中包含上下位关系的语句十分常见,同时,上下位关系是领域概念的一种基本关系。通过对文本中语句进行拆分,能够整理出建立本体所必须的概念以及概念之间的关系。

## 3 数控机床知识抽取及功能实现

### 3.1 知识抽取模型设计

结合领域本体的构建方法,Web 文本知识获取以及数控机床知识表示的相关理论和技术,提出一种基于领域本体的知识获取模型,如图 1 所示。

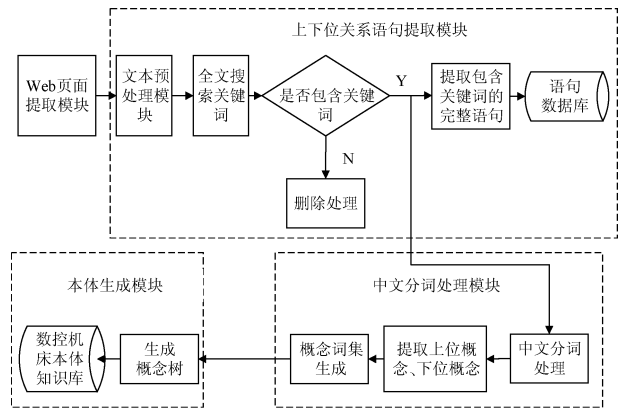


图 1 数控机床领域知识获取模型

Fig. 1 Knowledge acquisition model in the field of CNC machine tools

整个模型设计分为 4 个模块,分别是 Web 页面提取模块,上下位关系语句提取模块,中文分词处理模块,本体生成模块。

从 Web 页面中获得原始的 Web 文本资料需要 Web 页面抽取模块,该模块利用 C#语言,构建专用网络爬虫程序,遍历选定的高档数控机床相关网站,构建出该网站的页面树结构,利用宽度优先策略,以离初状态的状态距离为序依次对页面书上的节点遍历。

上下位关系语句提取模块将输入的原始电子文档进行预处理,统一格式后,对文档全文匹配表示上下位关系

的关键词,定位到关键词后,提取包含关键词的完整语句并储存到数据库中备用。

中文分词处理模块对提取的上下位关系语句进行中文分词切分。本文采用 ICTCLAS 将提取的上下位关系句子进行分解,通过定位到的关键词,获得对应的上位概念和下位概念,并保存到概念词汇数据库中。

本体生成模块将得到的概念词集生成概念树结构。将知识进行形式化表示,利用 Protégé 软件检验和调整生成的本体,完成知识获取。

### 3.2 上下位关系句子提取

本体概念的核心关系是上下位关系。其中,“是一种”模式是典型的上下位关系模式,它可以简单理解为,如果语句中包含“X 是一种 Y”,则 X 和 Y 具有上下位关系<sup>[10]</sup>。其中,特定性较强的单词叫做概括性较强的单词的下位词(hyponym),概括性较强的单词叫做特定性较强的单词的上位词(hyponym)<sup>[11]</sup>。例如在“加工中心是目前数控机床中产量最大、应用最广的数控机床。”这句话中,“加工中心”和“数控机床”作为两个概念,就存在上下位关系,“数控机床”是上位概念,“加工中心”是下位概念。在基于本体的知识表示中,概念之间的上下位关系主要是表示概念之间整体和部分的包含与被包含关系,继承关系,所属关系等。当前,上下位关系的获取方法主要有三种,分别是基于模式匹配的上下位关系获取,基于统计的上下位关系获取和基于网络百科的上下位关系获取<sup>[12]</sup>。本文采用基于模式匹配的上下位关系获取方法。

为了从原始 Web 文档中提取包含上下位关系的语句,需要将初始输入的电子文档标准化成统一格式的文本。具体内容包括去除 Web 文档中图片、公式等无法通过软件自动处理的元素和单元,去除空行等无意义的符号。最终输出为仅保留字符、标点符号、换行符且格式为 UTF-8 的标准化文本。

文本标准化之后进行“是一种”模式关键词匹配。在进行匹配过程中,关键词的选择是决定匹配正确率的关键词因素。关键词包括“是一种”,“是一类”,“是一个”,“是一部”等,可以概括为关键字“是一”加上一个量词的结构。将挑选的量词组成一个集合  $L$ ,  $L = (\text{种, 台, 个, 套, 名, 间, 件, 类, 支, 层, 笔, } \dots)$ , 进而组成第一类关键词集合。其他关键词还包括“可分为”,“由...组成”,“如下”,“包括”等包含类关键词,组成第二类型关键词。用这两种类型关键词建立关键词词典包。对经过预处理过的文本,与关键词词典进行匹配,如果检测到关键词,以关键词为起始位置,分别向前向后遍历,若遇到表示句子结束的标点符号则停止搜索,获得完整的语句。将同一 Web 页面提取到的语句存储到相应的数据库中。提取过程如图 2 所示。

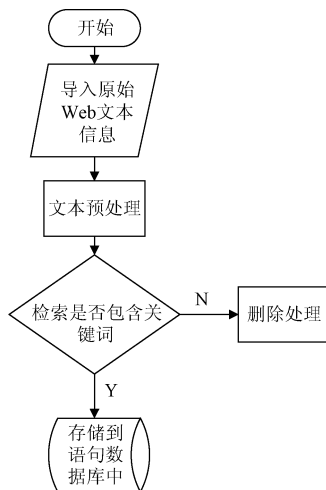


图 2 语句提取过程

Fig. 2 Sentence extraction process

### 3.3 中文分词处理

建立数控机床的领域本体,需要概念集合和概念树作为基础。作为产生概念集合的重要步骤,中文原子分词的准确性直接关系到领域本体的质量。目前,最具代表性的中文原子词切分系统有两种:

1) 由中国科学院计算技术研究所研发的基于统计方法和简单语法分析的 ICTCLAS 系统;

2) 由哈尔滨工业大学信息检索研究室研发的基于统计方法的 IRLAS 系统。这两个分词系统都能达到 90% 以上的准确率<sup>[13]</sup>。

本文采用 ICTCLAS 来进行语句的分词和原子词词性标注的操作。ICTCLAS 可以对一级词性,二级词性和三级词性中子类进行细致地划分。其中动词的二级词性划分为趋向动词、形式动词、不及物动词、动词是、动词有等。二级词性人名细分为汉语姓氏、汉语名字、音译人名等。经过 ICTCLAS 分词处理,可以得到生成概念集的素材。

3.2 节提取到的句子导入到程序中,进行原子切分,然后在此基础上,进行  $N$ -最短路径粗切分,比较切分结果,找出前  $N$  个最符合的切分结果,生成二元分词表,对得到的结果在进行基于类的隐马尔科夫分词,得到分词序列之后,对词进行词性标注。

为实现分词的划分和筛选,本文使用两层循环嵌套程序,第一层循环实现遍历所有原子功能,第二层循环实现将得到的原子进行验证的功能,将该原子与其后相邻的原子连接到一起,作为匹配对象,在词典数据库中进行匹配,验证该原子对是否为一个有意义的词,最终生成分词结果。

完成分词之后需要抽取语句中的上位概念和下位概念。由于 3.2 节中,采用匹配“是一种”模式关键词获取

到包含上下位关系的语句,所以得到的语句为简单句,即只包含一个主谓结构。

根据中文句子特点,首先确定关键词位置及类型,将关键词作为主谓结构中的谓语成分,获取上位概念和下位概念的问题就转换为定位主语成分和宾语成分的问题。根据关键词类型不同,抽取过程有两种情况,一种是匹配到“是一种”类型关键词的句子,上位概念即句子的宾语是从句子中“是一种”模式关键词右侧内容中定位名词,下位概念即句子的主语是从句子中“是一种”模式关键词左侧内容中定位名词。另一种是匹配到包含类关键词的句子,与第一种情况相反,关键词左侧名词为上位概念,右侧名词为下位概念。将上位概念和下位概念保存到概念数据库中,生成概念集,过程如图 3 所示。

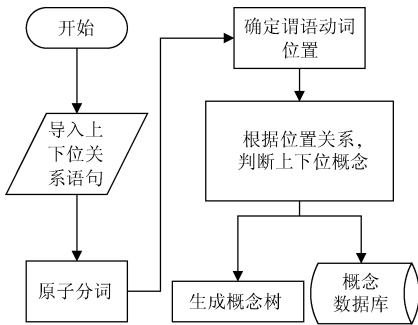


图 3 分词处理过程

Fig. 3 Processing of word segmentation

### 3.4 本体的检验及质量评价标准

为了让机器能够认识知识,必须用形式化的语言对本体进行编码。本文将抽取到的概念及概念之间的关系,通过程序直接用 OWL 语言表示。为了能够人为检验和修改生成的本体,采用美国斯坦福大学开发的基于知识的,开源的本体编辑工具 Protégé,对已经生成的本体文件进行编码操作。它拥有友好的界面风格,操作简便,提供了大量的知识模型架构与动作,用于创建、可视化、操纵各种表现形式的本体<sup>[14]</sup>。

本体质量评价指的是对自动或者手动构建的本体概念及概念之间的关系进行评价,来确定构建的领域本体是否能够准确地、完整地表达领域知识。评价指标分为绝对评价指标和对比评价指标两种。绝对评价中,以标准的专业术语集合为基础,通过采集的概念与标准术语比较,判定本体构建的质量。对比评价中,将人工建立的本体与自动建立的本体质量指标比较。童名文等《课程本体自动构建技术研究》<sup>[15]</sup>一文中,利用可译系数  $\alpha$ 、规模系数  $\beta$  和关系覆盖度作为指标对本体质量进行评价。本文利用这三种评价指标对提出的本体构建方法进行评定。

为描述方便,进行如下定义。

定义 1(可译系数  $\alpha$ ):

$$\alpha = M/N \tag{1}$$

式中: $M$  是抽取的概念与标准专业术语或概念匹配的个数, $N$  是抽取的概念总数。

定义 2(规模系数  $\beta$ ):

$$\beta = M/S \tag{2}$$

式中: $M$  是抽取的概念与标准专业术语或概念匹配的个数, $S$  是标准专业术语或概念总数。

定义 3(关系覆盖度  $rd$ ):

$$rd = R/S \tag{3}$$

式中: $R$  是构建的本体中,概念之间关系的总数, $N$  是抽取的概念总数。

## 4 实验及结果分析

### 4.1 构建本体

根据本文提出的本体获取方法,随机选取 50 个实验室机床设备知识介绍专用网页进行分析,利用上下位关系匹配的方式,进行 Web 页面的分析及概念获取,直接生成概念树和 owl 文件格式的本体表达,实验采用 ORACLE 数据库。

图 4 所示为模块对文本进行预处理,统一文本格式。

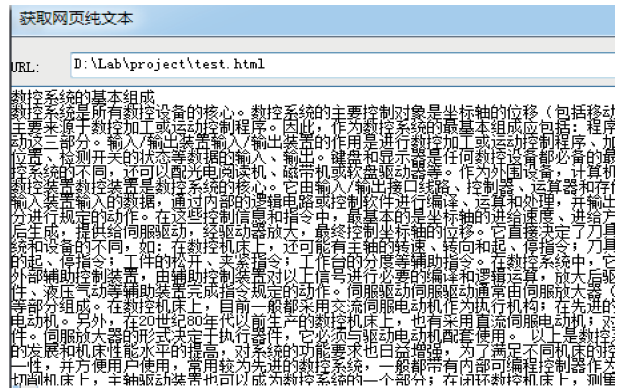


图 4 文本预处理

Fig. 4 Text preprocessing

实验中共提取了候选处理句子 2 336 句。其中包含标准术语共 153 个,程序抽取概念 87 个,关系 45 个。利用获取到的领域概念及其关系,生成如图 5 所示的概念树。这种概念树是组织概念和创建本体的基础。

将概念树中显示的概念及关系统一编码成 owl 文件,如下为 owl 文件片断。

```

<!-- #主传动系统 -->
<owl:Class rdf:about = "#主传动系统" >
< rdfs:subClassOf rdf:resource = "#机床机体" / >
</owl:Class >
<!-- #伺服放大器 -->
  
```

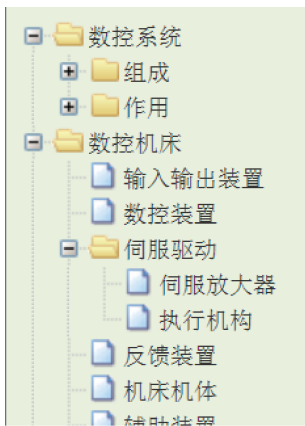


图5 本体概念树

Fig.5 Ontology concept tree

通过人工方式对 Web 页面进行上下位概念的获取,构建概念树,利用本体编辑工具 Protégé 对领域本体进行表示。在绝对评价分析中,采用第一组自动抽取概念的实验数据,如表 1 自动抽取概念评价表所示。

表 1 自动抽取概念评价

Table 1 Evaluation of automatically extracting concepts

标准术语数 (S)	抽取概念数 (N)	关系系数 (R)	匹配数 (M)	可译系数 (α)	规模系数 (β)	关系覆盖度 (rd)
153	87	53	77	0.885	0.503	0.609

在实验结果中,可译系数是抽取的概念中,与标准术语相符合的概念所占的比例,0.885 说明自动抽取的概念准确度较高,抽取结果可以信赖。规模系数反应的是抽取正确概念的能力,由于本文提出的本体构建方法是利用上下位关系匹配的方法进行概念抽取,所以关键词词典包词汇量的多少直接决定了抽取概念的数量,合理增加关键词,能够提高概念的抽取能力,得到更加完整,知识颗粒度更小的数控机床领域本体。关系覆盖度体现的是概念与概念之间关联程度,即关系覆盖度越小,概念之间关系越复杂。由定义 3 计算得出实验网页文本中,实际的关系覆盖度为 0.582,对比实验结果中关系覆盖度 0.609 可以发现,实验结果与实际值非常接近,说明实验得到的本体能够有效的对原有知识进行表示。

在相对评价分析中,将第 1 组自动抽取概念和第 2 组人工抽取概念实验数据进行对比,如表 2 对比试验数据表所示。

表 2 对比试验数据

Table 2 Contrast test data

获取方式	抽取概念数 (N)	关系系数 (R)	匹配数 (M)	可译系数 (α)	关系覆盖度 (rd)	消耗时间 (h)
自动	87	45	77	0.885	0.517	0.3
人工	118	67	106	0.898	0.576	>20h

根据表中实验数据分析,人工抽取概念试验中,可译系数和关系覆盖度与自动抽取概念实验得到的数据十分接近,从准确度和表达概念关系复杂程度来看,两种方法没有很大差别,但是,人工抽取概念花费的时间要远远大于自动抽取概念花费的时间,在构建本体的效率上,本文提出的基于本体的数控机床知识获取方法具有相当大的优势,而且,为了保证软件稳定性,软件编写中只运用单线程处理数据,如果对软件进行优化,采用多线程处理数据,将会进一步减少抽取概念和构建本体所消耗的时间。

## 5 结论

本文提出一种基于本体的数控机床知识获取模型,

```

<owl:Class rdf:about = "#伺服放大器" >
<rdfs:subClassOf rdf:resource = "#伺服驱动" / >
</owl:Class >
<! -- #伺服驱动 -- >
<owl:Class rdf:about = "#伺服驱动" >
<rdfs:subClassOf rdf:resource = "#数控系统" / >
</owl:Class >
<! -- #作用 -- >
<owl:Class rdf:about = "#作用" >
<rdfs:subClassOf rdf:resource = "#数控机床" / >
</owl:Class >
    
```

通过上述概念的抽取及 OWL 语言表示,实现了本体的构建和相关知识的表示。为了直观的检验创建本体的知识表达的有效性,将 OWL 文件导入到 Protégé 工具中,图 6 所示为软件中显示本体中类的列表及类的层次图。

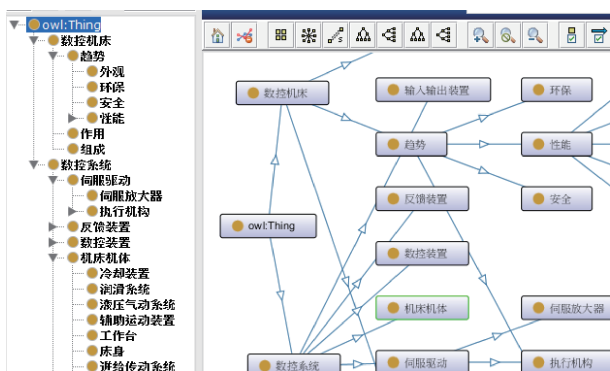


图 6 类列表及层次图

Fig.6 Class list and hierarchical graph

## 4.2 本体质量评价

为了直观地验证方法的有效性和可行性,本文设计了两组对比实验,第 1 组是 4.1 中构建的本体;第 2 组是

利用网络爬虫程序,获得原始 Web 页面信息,预处理后的 Web 文本经过提取上下位关系语句和中文分词,得到了概念集,以此为基础构建了数控机床知识领域本体,完成了知识获取。该方法有效地屏蔽了语义相关度的计算,提高了知识库的构建效率。但是本文提出的方法对匹配的关键词字典依赖程度较高,建立的数控机床领域本体具有局限性,并且由于数控机床领域知识数量庞大,依靠 Web 文本只能获取一部分知识,今后还需要对其他数控机床知识获取的方式做进一步研究。

**参考文献**

[ 1 ] 陈钊,张冬梅. Web 信息抽取技术综述[J]. 计算机应用研究,2010, 27 ( 12 ):4401-4405.  
CHEN ZH, ZHANG D M. Overview of Web information extraction technology [ J ]. Application Research of Computers, 2010, 27 ( 12 ): 4401-4405.

[ 2 ] 生佳根. 基于本体的知识获取、管理和应用方法研究[D]. 南京:南京航空航天大学,2012.  
SHENG J G. Research on knowledge acquisition, management and application based on ontology [ D ]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2012.

[ 3 ] 邓小亚. 本体构建研究综述[J]. 现代计算机:专业版,2012(5):19-22.  
DENG X Y. Overview of ontology construction [ J ]. Modern Computer: Professional Edition, 2012 ( 5 ): 19-22.

[ 4 ] 田俊华. 基于本体知识库的教学资源自动采集技术研究[D]. 南京:南京师范大学,2011.  
TIAN J H. Research on automatic collection of teaching resources based on ontology knowledge database [ D ]. Nanjing: Nanjing Normal University, 2011.

[ 5 ] 冉婕,漆丽娟. 本体的形式化研究[J]. 微型机与应用,2012,31 ( 6 ):1-3.  
RAN J, QI L J. Research on formalization of ontology [ J ]. Microcomputer & Its Applications, 2012, 31 ( 6 ): 1-3.

[ 6 ] 李长杰,明新国,邱坤华,等. 基于本体的飞机工装设计知识表示方法[J]. 中国机械工程,2014, 25 ( 19 ): 2614-2619.  
LI CH J, MING X G, QIU K H, et al. The representation method of aircraft tooling design based on ontology [ J ]. China Mechanical Engineering, 2014, 25(19): 2614-2619.

[ 7 ] 刘建伟,燕路峰. 知识表示方法比较[J]. 计算机系统应用,2011, 20 ( 3 ):242-246.  
LIU J W, YAN F F. Comparison of knowledge representation methods [ J ]. Computer Systems & Applications, 2011, 20 ( 3 ): 242-246.

[ 8 ] 王昌飞,徐海波. 面向对象混合型知识表示在冰箱设计中的应用[J]. 机械工程与自动化,2009(2):2-4.  
WANG CH F, XU H B. Application of object oriented

hybrid knowledge representation in refrigerator design [ J ]. Mechanical Engineering & Automation, 2009 ( 2 ): 2-4.

[ 9 ] 曲朝阳,孙鹏飞. 基于本体语言 OWL 的电网领域知识表示方法 [ J ]. 东北电力大学学报,2012, 32 ( 4 ): 30-34.  
QU CH Y, SUN P F. A knowledge representation method based on ontology language OWL in power grid [ J ]. Journal of Northeast Dianli University, 2012, 32 ( 4 ): 30-34.

[ 10 ] 蒋年树. 领域本体概念上下位关系的抽取及组织研究[D]. 昆明:昆明理工大学,2013.  
JIANG N SH. Research on the extraction and organization of domain ontology concepts [ D ]. Kunming : Kunming University of Science and Technology, 2013.

[ 11 ] 奉国和,郑伟. 国内中文自动分词技术研究综述[J]. 图书情报工作,2011, 55 ( 2 ):41-45.  
FENG G H, ZHENG W. A summary of the research on Chinese automatic word segmentation in China [ J ]. Library and Information Service, 2011, 55 ( 2 ):41-45.

[ 12 ] 刘磊,曹存根,张春霞,等. 概念空间中上下位关系的意义识别研究 [ J ]. 计算机学报,2009, 32 ( 8 ): 1651-1661.  
LIU L, CAO C G, ZHANG CH X, et al. Research on the meaning recognition of hyponymy in concept space [ J ]. Chinese Journal of Computers, 2009, 32 ( 8 ): 1651-1661.

[ 13 ] 何月顺. 关联规则挖掘技术的研究及应用 [ D ]. 南京:南京航空航天大学,2010.  
HE Y SH. Research and application of mining association rules [ D ]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2010.

[ 14 ] 张文秀,朱庆华. 领域本体的构建方法研究 [ J ]. 图书与情报,2011(1):16-19,40.  
ZHANG W X, ZHU Q H. Research on domain ontology construction method [ J ]. Library and Information, 2011(1): 16-19, 40.

[ 15 ] 童名文,牛琳,杨琳,等. 课程本体自动构建技术研究 [ J ]. 计算机科学,2016, 43 ( S2 ):108-112.  
TONG M W, NIU L, YANG L, et al. Research on the automatic construction of curriculum ontology [ J ]. Computer Science, 2016, 43 ( S2 ): 108-112.

**作者简介**



刘锴锋,1990 年出生,2009 年于北京信息科技大学获得学士学位,现为北京信息科技大学在读硕士研究生,主要研究方向为机械设计制造及其自动化。  
E-mail:1029370027@qq.com

**Liu Kaifeng** was born in 1990, M. Sc. candidate in Beijing Information Science & Technology University. His main research direction is mechanism design, manufacturing and automatization.