



(12) 发明专利申请

(10) 申请公布号 CN 113987212 A

(43) 申请公布日 2022. 01. 28

(21) 申请号 202111361153.5

G06N 3/04 (2006.01)

(22) 申请日 2021.11.17

G06F 16/901 (2019.01)

G06F 16/904 (2019.01)

(71) 申请人 武汉理工大学

地址 430000 湖北省武汉市洪山区珞南街
道珞狮路205号

(72) 发明人 萧箏 王继业 田野

(74) 专利代理机构 武汉市首臻知识产权代理有
限公司 42229

代理人 朱迪

(51) Int. Cl.

G06F 16/36 (2019.01)

G06F 16/34 (2019.01)

G06F 16/35 (2019.01)

G06F 40/289 (2020.01)

G06F 40/30 (2020.01)

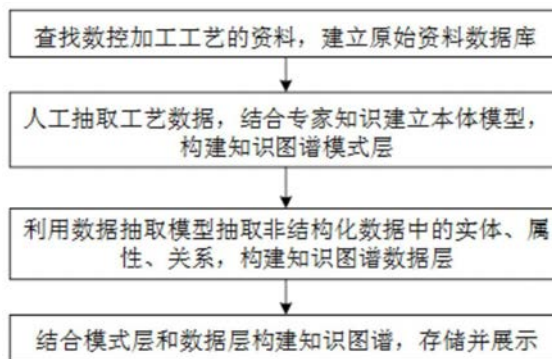
权利要求书5页 说明书15页 附图5页

(54) 发明名称

一种数控加工领域工艺数据的知识图谱构建方法

(57) 摘要

一种数控加工领域工艺数据的知识图谱构建方法,包括以下步骤:S1:建立原始资料数据库,S2:构建知识图谱模式层,人工抽取原始资料数据库中的工艺知识,结合数控加工工艺领域的专家知识建立本体模型,构建知识图谱模式层;S3:构建知识图谱数据层、提取三元组,利用数据抽取模型抽取非结构化数据中的实体、属性、关系,构建知识图谱数据层;S4:可视化展示,模式层和数据层结合起来构建知识图谱,存储知识图谱并进行可视化展示。本设计不仅能够直观、全面地管理工艺数据,而且在基于深度学习推理的前提下实现数据的深层开发。



1. 一种数控加工领域工艺数据的知识图谱构建方法,其特征在于:

所述工艺数据的知识图谱构建方法包括以下步骤:

S1: 建立原始资料数据库,查找数控加工领域相关的数据资料,并建立原始资料数据库;

S2: 构建知识图谱模式层,人工抽取原始资料数据库中的工艺知识,结合数控加工工艺领域的专家知识建立本体模型,构建知识图谱模式层;

S3: 构建知识图谱数据层、提取三元组,利用数据抽取模型抽取非结构化数据中的实体、属性、关系,构建知识图谱数据层;由于原始资料数据库中的数据均属于包含大量信息的非结构化工艺数据,要完成知识图谱的自动建立,则需要将包含大量信息的非结构化工艺数据的信息进行抽取,将非结构化数据转化为语义关系数据;故建立数据抽取模型自动抽取非结构化数据中的实体、属性、关系,构建知识图谱数据层;所述数据抽取模型包括工艺语句分词模块、工艺词语分类模块和语义关系构建模块;

S4: 可视化展示,模式层和数据层结合起来构建知识图谱,存储知识图谱并进行可视化展示。

2. 根据权利要求1所述的一种数控加工领域工艺数据的知识图谱构建方法,其特征在于:

所述S1: 建立原始资料数据库中的数控加工领域相关的数据资料包括:零件的图纸数据、规范性手册、加工过程工艺卡片、数控机床和加工中心的数据、夹具量具定位装备的数据、工艺文件、工序卡片、专业书籍、工艺手册;将查找到的数控加工领域相关的数据资料,存储到本地硬盘中作为原始资料数据库。

3. 根据权利要求2所述的一种数控加工领域工艺数据的知识图谱构建方法,其特征在于:

所述S2: 构建知识图谱模式层,首先人工抽取原始资料数据库中的工艺知识,结合数控加工工艺领域的专家知识建立本体模型,构建知识图谱模式层;

S2.1: 分析抽取出来的工艺数据,结合专家知识对数据进行归纳;要建立本体模型并构建知识图谱模式层;知识图谱模式层通过本体模型方式来构建,本体模型通过本体语言OWL来描述;针对原始资料数据中的非结构化数据通过人工提取工艺知识,然后结合专家知识建立本体模型构建知识图谱模式层;本体模型是模式层的模型,本体语言是描述模式层的语言;

S2.2: 数控加工设备的加工对象是零部件,分析零部件的CAD模型和数控加工工艺文档包含的零部件信息及加工制造信息,可以将其归纳为六类数据:加工对象数据、加工设备数据、加工方法数据、加工装备数据、加工特征数据以及语义关系数据;

工艺数据可以总结为:加工对象数据、加工设备数据、加工方法数据、加工装备数据、加工特征数据和语义关系数据;

加工对象数据是整个工艺数据的基础,描述了加工对象的编号、名称等文本信息,还描述了包括机械模型的材料、尺寸、要求的信息;加工设备数据主要指加工零件需要的数控机床或者加工中心,例如数控铣床、数控车床;加工方法数据是指加工零件的方法和操作,材料从毛坯到零件要经过的步骤,具体指加工工序和工步;加工装备数据指的是工艺装备,包括定位用到的定位装置、夹紧装置,加工过程中的量具辅具,如机床上的刀具;加工特征数

据是指机床所完成的工序中具体的加工对象及要求,如面、孔、螺纹,加工要求包括精度要求和材料处理的要求;语义关系数据是指前面几种数据之间的关系,表示不同类别的数据之间的联系;

S2.3:为了保持模式层的准确性和完备性,把以上的六种数据抽象为六大类,其中加工对象数据、加工设备数据、加工方法数据、加工装备数据、加工特征数据以及语义关系数据依次对应:零件类、设备类、工艺类、装备类、特征类、关系类;工艺类继承有工序类和工步类,装备类继承有刀具类、夹具类、量具类、辅具类,特征类继承有形状特征类、精度特征类、技术特征类、管理特征类、材料特征类,其中,精度特征类继承有尺寸精度类、形状精度类、位置精度类、表面粗糙度类,关系类继承有contain类、order类;

类的属性包括类属性和实例属性,类属性为各类及其子类共享,所有实例共享对应的类属性,实例属性仅为各个实例所有;零件类具有名称类属性,设备类具有名称类属性,工艺类具有名称类属性,装备类具有名称类属性,特征类具有名称类属性,关系类具有名称类属性,其他实例属性根据不同实体添加;

S2.4:知识图谱的模式层描述一些实体类及之间的关系,模式层根据S2.3中抽象的实体类建立,通过表达式来描述,把抽象出来的实体类和关系建立本体模型,利用protégé软件建立和展示,模式层的知识通过本体模型来表示,本体模型可以用owl语言来描述;

知识图谱的模式层的本体模型形式化表达为:

$$KGPattern = \{Entity \cup Relation\}$$

其中:

$$Entity = \{P \cup M \cup O \cup E \cup F\}$$

$$Relation = \{R\}$$

$$R = \{(P_i, contain, F_j) \cup (O_i, contain, M_j) \cup (O_i, contain, O_j) \cup (O_i, contain, E_j) \cup (O_i, contain, F_j) \cup (O_i, order, O_j) \cup \}$$

以上式子中:KGPattern指知识图谱模式层的形式化表达模型,Entity指模式层描述的实体集合,Relation指模式层描述的关系集合;P表示零件类,M表示设备类,O表示工艺类,E表示装备类,F表示特征类,R表示关系类;语义关系R包括contain、order,contain关系定义为零件类与特征类之间的包含关系,工艺类与设备类、装备类、特征类的包含关系,以及工艺类之间的包含关系,关系表示及描述为:($P_i, contain, F_j$)表示零件 P_i 包含特征 F_j , ($O_i, contain, M_j$)表示工艺 O_i 包含设备 M_j , ($O_i, contain, O_j$)表示工艺 O_i 包含工艺 O_j , ($O_i, contain, E_j$)表示工艺 O_i 包含装备 E_j , ($O_i, contain, F_j$)表示工艺 O_i 包含特征 F_j ;order关系定义为工艺类中工序之间或工步之间的顺序关系,关系表示及描述为:($O_i, order, O_j$)表示工序(工步) O_i 在前,工序(工步) O_j 在后;如表1所示;

表1语义关系表示及描述

语义关系	关系类表示	关系类描述
<i>contain</i>	$(P_i, contain, F_j)$	零件 P_i 包含特征 F_j
	$(O_i, contain, M_j)$	工艺 O_i 包含设备 M_j
	$(O_i, contain, O_j)$	工艺 O_i 包含工艺 O_j
	$(O_i, contain, E_j)$	工艺 O_i 包含装备 E_j
	$(O_i, contain, F_j)$	工艺 O_i 包含特征 F_j
<i>order</i>	$(O_i, order, O_j)$	工序 O_i 在前, 工序 O_j 在后或工步 O_i 在前, 工步 O_j 在后

此时,知识图谱模式层构建完成,进入下一步S3。

4. 根据权利要求3所述的一种数控加工领域工艺数据的知识图谱构建方法,其特征在于:

所述S3:构建知识图谱数据层、提取三元组,利用数据抽取模型抽取非结构化数据中的实体、属性、关系,构建知识图谱数据层;

由于原始资料数据库中的数据均属于包含大量信息的非结构化工艺数据,要完成知识图谱的自动建立,则需要将包含大量信息的非结构化工艺数据的信息进行抽取,将非结构化数据转化为语义关系数据;故建立数据抽取模型自动抽取非结构化数据中的实体、属性、关系,构建知识图谱数据层;所述数据抽取模型包括工艺语句分词模块、工艺词语分类模块和语义关系构建模块;具体运行流程如下:

S3.1:把待处理的非结构化数据转化为纯文本格式文件,通过编程读入数据,把数据保存为txt格式文件;

S3.2:利用工艺语句分词模块分词,所述工艺语句分词模块基于Bi-LSTM-CRF算法实现,利用的工艺语句分词模块分词,把S3.1中得到的纯文本数据输入到训练好的工艺语句分词模块中,最终输出带有分词标签“B”、“I”、“W”、“S”的文本文件,此时文本分词完成:

S3.3:将分词完成的文本内的词语进行分类形成词向量,工艺词语分类模块基于word2vec算法实现,将S3.2得到的带有分词标签的文本文件中的分词作为待分类词语,计算待分类词语与加工词语分类语料库的词语的余弦距离,将距离最短的样本的类别赋予待分类词语相同的类别,最终输出带有词语类别标签的文本文件,此时词语分类完成;

S3.4:语义关系构建,将S3.3中得到的带有词语类别标签的文本文件,根据模式层定义的实体类别来识别词语,根据模式层定义的实体之间的关系来识别词语之间的关系,并以三元组方式存储,根据词语属性标签来识别实体和属性之间的关系,也以三元组方式保存,

保存的三元组输出为<实体,关系,实体>,<实体,属性,属性值>三元组。

5. 根据权利要求4所述的一种数控加工领域工艺数据的知识图谱构建方法,其特征在于:

所述S3.4:语义关系构建,将S3.3中得到的带有词语类别标签的文本剔除无类别的占位词,得到有类别的词语文本集合,按顺序依次处理类别的词语文本集合内的词语得到三元组输出,直到所有词语处理完成;所述处理过程如下:

S3.4.1:如前者词语标签为“p”,且下一个词语标签为“f”,则添加三元组<前者词语,contain,后者词语>,然后进入下一词语的处理;如果不是则进入S3.4.2;

S3.4.2:如前者词语标签为“p、m、e、f、d、s”中的任意一个,且下一个词语标签为“a”,则添加三元组<前者词语,ofattribute,后者词语>,然后进入下一词语的处理;如果不是则进入S3.4.3;

S3.4.3:如前者词语标签为“d”或“s”,且下一个词语标签为“m”、“e”或“f”,则添加三元组<前者词语,contain,后者词语>,然后进入下一词语的处理;如果不是则进入S3.4.4;

S3.4.4:如前者词语标签为“d”,且下一个词语标签为“s”,则添加三元组<前者词语,contain,后者词语>,然后进入下一词语的处理;如果不是则进入S3.4.4;

S3.4.5:如前者词语标签和下一个词语标签均为“d”,或前者词语标签和下一个词语标签均为“s”,则添加三元组<前者词语,order,后者词语>,然后进入下一词语的处理;如果不是则直接进入下一词语的处理;

所有词语处理完成后将其以csv文件的格式保存,此时语义关系构建完成,进入S4。

6. 根据权利要求4所述的一种数控加工领域工艺数据的知识图谱构建方法,其特征在于:

所述S3.2中的工艺语句分词模块的训练过程如下:

S3.2.1:利用分词包jieba对纯文本格式文件中的语句进行初步分词;

S3.2.2:人工对S3.2.1中得到的分词结果进行判断修改,标注“B”、“I”、“W”、“S”标签,其中“B”表示词的首汉字、“I”表示词的中间汉字、“W”表示词的末尾汉字、“S”表示词仅有一个汉字;

S3.2.3:将S3.2.2中得到的带有分词标签的文本存储到本地硬盘中作为工艺语句分词语料库;

S3.2.4:训练工艺语句分词模块,将S3.2.3得到的工艺语句分词语料库导入到工艺语句分词模块,然后读入分词模块程序进行迭代运行,程序运行完成后工艺语句分词模块训练完成。

7. 根据权利要求4所述的一种数控加工领域工艺数据的知识图谱构建方法,其特征在于:

所述S3.3中工艺词语分类模块的训练过程如下:

S3.3.1:人工对上面处理好的文本中的词语标注“零件”、“设备”、“装备”、“特征”、“关系”、“工序”、“工步”、“属性”八个类别,分别用“p”、“m”、“e”、“f”、“r”、“d”、“s”、“a”表示,无类别词用“x”表示,上述“零件”、“设备”、“装备”、“特征”、“关系”、“工序”、“工步”七个类别与S2.2中的六类数据相对应,“属性”与S2.3中的类的属性相对应;

加工对象数据对应“零件”;加工设备数据对应“设备”;加工方法数据对应“工序”、“工

步”，零件加工包含多个工序，每个工序包含多个工步；加工装备数据对应“装备”；加工特征数据对应“特征”；语义关系数据对应“关系”；“属性”包含数值和名称；

S3.3.2: 将S3.3.1中得到的带有分类标签的文本存储到本地硬盘中作为工艺词语分类语料库；

S3.3.3: 训练工艺词语分类模块，将S3.3.2得到的工艺词语分类语料库导入到工艺词语分类模块，然后读入分类模块程序进行迭代运行，程序运行完成后工艺词语分类模块训练完成。

8. 根据权利要求4-7中任意一项所述的一种数控加工领域工艺数据的知识图谱构建方法，其特征在于：

S4: 可视化展示，把模式层和数据层结合起来构建知识图谱，把抽取到的三元组输入到Neo4j图数据库中，存储知识图谱并可视化展示。

一种数控加工领域工艺数据的知识图谱构建方法

技术领域

[0001] 本发明涉及一种数控加工领域工艺数据的知识图谱构建方法,具体适用于直观管理工艺数据。

背景技术

[0002] 数控加工以数控机床或数控加工中心为独立单位进行加工制造,数控机床针对单个零件加工制造,毛坯材料流转在不同的机床最终加工成型。工艺数据指对机械产品进行工艺设计过程中使用的和所产生的数据,包括规范化的静态数据和加工过程中产生的动态数据。从零件的设计到加工合格会产生大量的数据,数据类型复杂,数据格式多样,结构化数据、半结构化数据以及非结构化数据都有包含。零件加工过程中的动态数据由数控加工设备搜集并存储在数据库中,零件设计过程涉及到的标准文件多是非结构化数据,零件的设计图纸格式也因建模软件的不同而不同,零件的加工工艺根据不同的加工情况和企业也有不同的呈现方式,比如工序卡片的格式不是完全相同的,加工流程是否要以工序卡片来表达也是不同的。从工艺数据内容来看,可分为资源类数据、规则类数据、规程类数据。结构化数据以及半结构化数据与知识图谱关联比较容易,而非结构化数据用知识图谱来表示较难实现。非结构化数据的表达内容丰富,但是进一步挖掘利用数据的价值很难实现。所以提出了数控加工领域工艺数据中的非结构化数据的知识图谱构建方法。

[0003] 知识图谱是结构化表示知识的一种方法,它以图的形式来描述客观事实及它们之间的联系。从本质上来看,知识图谱是一种描述实体、实体属性与实体关系的大型语义网络,其中的节点可以代表实体和属性,边可以代表关系。主要采用三元组的形式进行表示,只要把<实体,关系,实体>、<实体,属性,属性值>所有三元组描述清楚知识图谱就描述清楚了。所以,知识图谱构建的重点也是三元组信息的抽取。知识图谱可分为通用知识图谱和垂直领域知识图谱,垂直领域知识图谱专业性更强,精度较高,知识体系更加完善。知识图谱具有直观、学习性强等优点。

[0004] 数控加工领域的工艺数据复杂多样,有的研究是基于BOM来管理和集成工艺数据的,有的研究是基于EBOP工艺结构树来组织和管理工艺数据的,有的研究是通过开发工艺数据库系统来存储和利用工艺数据的。这些对工艺数据的管理办法不够直观,对知识的利用性不高,难以挖掘数据的深层次价值。

发明内容

[0005] 本发明的目的是克服现有技术中存在的工艺数据管理不够直观全面的问题,提供了一种直观全面管理工艺数据的数控加工领域工艺数据的知识图谱构建方法。

[0006] 为实现以上目的,本发明的技术解决方案是:

[0007] 一种数控加工领域工艺数据的知识图谱构建方法,包括以下步骤:

[0008] S1:建立原始资料数据库,查找数控加工领域相关的数据资料,并建立原始资料数据库;

[0009] S2:构建知识图谱模式层,人工抽取原始资料数据库中的工艺知识,结合数控加工工艺领域的专家知识建立本体模型,构建知识图谱模式层;

[0010] S3:构建知识图谱数据层、提取三元组,利用数据抽取模型抽取非结构化数据中的实体、属性、关系,构建知识图谱数据层;由于原始资料数据库中的数据均属于包含大量信息的非结构化工艺数据,要完成知识图谱的自动建立,则需要将包含大量信息的非结构化工艺数据的信息进行抽取,将非结构化数据转化为语义关系数据;故建立数据抽取模型自动抽取非结构化数据中的实体、属性、关系,构建知识图谱数据层;所述数据抽取模型包括工艺语句分词模块、工艺词语分类模块和语义关系构建模块;

[0011] S4:可视化展示,模式层和数据层结合起来构建知识图谱,存储知识图谱并进行可视化展示。

[0012] 所述S1:建立原始资料数据库中的数控加工领域相关的数据资料包括:零件的图纸数据、规范性手册、加工过程工艺卡片、数控机床和加工中心的数据、夹具量具定位装备的数据、工艺文件、工序卡片、专业书籍、工艺手册;将查找到的数控加工领域相关的数据资料,存储到本地硬盘中作为原始资料数据库。

[0013] 所述S2:构建知识图谱模式层,首先人工抽取原始资料数据库中的工艺知识,结合数控加工工艺领域的专家知识建立本体模型,构建知识图谱模式层;

[0014] S2.1:分析抽取出来的工艺数据,结合专家知识对数据进行归纳;要建立本体模型并构建知识图谱模式层;知识图谱模式层通过本体模型方式来构建,本体模型通过本体语言OWL来描述;针对原始资料数据中的非结构化数据通过人工提取工艺知识,然后结合专家知识建立本体模型构建知识图谱模式层;本体模型是模式层的模型,本体语言是描述模式层的语言;

[0015] S2.2:数控加工设备的加工对象是零部件,分析零部件的CAD模型和数控加工工艺文档包含的零部件信息及加工制造信息,可以将其归纳为六类数据:加工对象数据、加工设备数据、加工方法数据、加工装备数据、加工特征数据以及语义关系数据;

[0016] 工艺数据可以总结为:加工对象数据、加工设备数据、加工方法数据、加工装备数据、加工特征数据和语义关系数据;

[0017] 加工对象数据是整个工艺数据的基础,描述了加工对象的编号、名称等文本信息,还描述了包括机械模型的材料、尺寸、要求的信息;加工设备数据主要指加工零件需要的数控机床或者加工中心,例如数控铣床、数控车床;加工方法数据是指加工零件的方法和操作,材料从毛坯到零件要经过的步骤,具体指加工工序和工步;加工装备数据指的是工艺装备,包括定位用到的定位装置、夹紧装置,加工过程中的量具辅具,如机床上的刀具;加工特征数据是指机床所完成的工序中具体的加工对象及要求,如面、孔、螺纹,加工要求包括精度要求和材料处理的要求;语义关系数据是指前面几种数据之间的关系,表示不同类别的数据之间的联系;

[0018] S2.3:为了保持模式层的准确性和完备性,把以上的六种数据抽象为六大类,其中加工对象数据、加工设备数据、加工方法数据、加工装备数据、加工特征数据以及语义关系数据依次对应:零件类、设备类、工艺类、装备类、特征类、关系类;工艺类继承有工序类和工步类,装备类继承有刀具类、夹具类、量具类、辅具类,特征类继承有形状特征类、精度特征类、技术特征类、管理特征类、材料特征类,其中,精度特征类继承有尺寸精度类、形状精度

类、位置精度类、表面粗糙度类,关系类继承有contain类、order类;

[0019] 类的属性包括类属性和实例属性,类属性为各类及其子类共享,所有实例共享对应的类属性,实例属性仅为各个实例所有;零件类具有名称类属性,设备类具有名称类属性,工艺类具有名称类属性,装备类具有名称类属性,特征类具有名称类属性,关系类具有名称类属性,其他实例属性根据不同实体添加;

[0020] S2.4:知识图谱的模式层描述一些实体类及之间的关系,模式层根据S 2.3中抽象的实体类建立,通过表达式来描述,把抽象出来的实体类和关系建立本体模型,利用protégé软件建立和展示,模式层的知识通过本体模型来表示,本体模型可以用owl语言来描述;

[0021] 知识图谱的模式层的本体模型形式化表达为:

[0022] $KGPattern = \{Entity \cup Relation\}$

[0023] 其中:

[0024] $Entity = \{P \cup M \cup O \cup E \cup F\}$

[0025] $Relation = \{R\}$

[0026] $R = \{(P_i, contain, F_j) \cup (O_i, contain, M_j) \cup (O_i, contain, O_j) \cup (O_i, contain, E_j) \cup (O_i, contain, F_j) \cup (O_i, order, O_j) \cup \}$

[0027] 以上式子中:KGPattern指知识图谱模式层的形式化表达模型,Entity指模式层描述的实体集合,Relation指模式层描述的关系集合;P表示零件类,M表示设备类,O表示工艺类,E表示装备类,F表示特征类,R表示关系类。语义关系R包括contain、order,contain关系定义为零件类与特征类之间的包含关系,工艺类与设备类、装备类、特征类的包含关系,以及工艺类之间的包含关系,关系表示及描述为:($P_i, contain, F_j$)表示零件 P_i 包含特征 F_j ,($O_i, contain, M_j$)表示工艺 O_i 包含设备 M_j ,($O_i, contain, O_j$)表示工艺 O_i 包含工艺 O_j ,($O_i, contain, E_j$)表示工艺 O_i 包含装备 E_j ,($O_i, contain, F_j$)表示工艺 O_i 包含特征 F_j ;order关系定义为工艺类中工序之间或工步之间的顺序关系,关系表示及描述为:($O_i, order, O_j$)表示工序(工步) O_i 在前,工序(工步) O_j 在后;如表1所示;

[0028] 表1语义关系表示及描述

语义关系	关系类表示	关系类描述
[0029] <i>contain</i>	$(P_i, contain, F_j)$	零件 P_i 包含特征 F_j
	$(O_i, contain, M_j)$	工艺 O_i 包含设备 M_j
	$(O_i, contain, O_j)$	工艺 O_i 包含工艺 O_j
	$(O_i, contain, E_j)$	工艺 O_i 包含装备 E_j
	$(O_i, contain, F_j)$	工艺 O_i 包含特征 F_j

[0030]	<i>order</i>	$(O_i, order, O_j)$	工序 O_i 在前, 工序 O_j 在后或工步 O_i 在前, 工步 O_j 在后
--------	--------------	---------------------	---

[0031] 此时,知识图谱模式层构建完成,进入下一步S3。

[0032] 所述S3:构建知识图谱数据层、提取三元组,利用数据抽取模型抽取非结构化数据中的实体、属性、关系,构建知识图谱数据层;

[0033] 由于原始资料数据库中的数据均属于包含大量信息的非结构化工艺数据,要完成知识图谱的自动建立,则需要将包含大量信息的非结构化工艺数据的信息进行抽取,将非结构化数据转化为语义关系数据;故建立数据抽取模型自动抽取非结构化数据中的实体、属性、关系,构建知识图谱数据层;所述数据抽取模型包括工艺语句分词模块、工艺词语分类模块和语义关系构建模块;具体运行流程如下:

[0034] S3.1:把待处理的非结构化数据转化为纯文本格式文件,通过编程读入数据,把数据保存为txt格式文件;

[0035] S3.2:利用工艺语句分词模块分词,所述工艺语句分词模块基于Bi-LSTM-CRF算法实现,利用的工艺语句分词模块分词,把S3.1中得到的纯文本数据输入到训练好的工艺语句分词模块中,最终输出带有分词标签“B”、“I”、“W”、“S”的文本文件,此时文本分词完成:

[0036] S3.3:将分词完成的文本内的词语进行分类形成词向量,工艺词语分类模块基于word2vec算法实现,将S3.2得到的带有分词标签的文本文件中的分词作为待分类词语,计算待分类词语与加工词语分类语料库的词语的余弦距离,将距离最短的样本的类别赋予待分类词语相同的类别,最终输出带有词语类别标签的文本文件,此时词语分类完成;

[0037] S3.4:语义关系构建,将S3.3中得到的带有词语类别标签的文本文件,根据模式层定义的实体类别来识别词语,根据模式层定义的实体之间的关系来识别词语之间的关系,并以三元组方式存储,根据词语属性标签来识别实体和属性之间的关系,也以三元组方式保存,保存的三元组输出为<实体,关系,实体>,<实体,属性,属性值>三元组。

[0038] 所述S3.4:语义关系构建,将S3.3中得到的带有词语类别标签的文本剔除无类别的占位词,得到有类别的词语文本集合,按顺序依次处理类别的词语文本集合内的词语得到三元组输出,直到所有词语处理完成;所述处理过程如下:

[0039] S3.4.1:如前者词语标签为“p”,且下一个词语标签为“f”,则添加三元组<前者词语,contain,后者词语>,然后进入下一词语的处理;如果不是则进入S3.4.2;

[0040] S3.4.2:如前者词语标签为“p、m、e、f、d、s”中的任意一个,且下一个词语标签为“a”,则添加三元组<前者词语,ofattribute,后者词语>,然后进入下一词语的处理;如果不是则进入S3.4.3;

[0041] S3.4.3:如前者词语标签为“d”或“s”,且下一个词语标签为“m”、“e”或“f”,则添加三元组<前者词语,contain,后者词语>,然后进入下一词语的处理;如果不是则进入S3.4.4;

[0042] S3.4.4:如前者词语标签为“d”,且下一个词语标签为“s”,则添加三元组<前者词

语,contain,后者词语>,然后进入下一词语的处理;如果不是则进入S3.4.4;

[0043] S3.4.5:如前者词语标签和下一个词语标签均为“d”,或前者词语标签和下一个词语标签均为“s”,则添加三元组<前者词语,order,后者词语>,然后进入下一词语的处理;如果不是则直接进入下一词语的处理;

[0044] 所有词语处理完成后将其以csv文件的格式保存,此时语义关系构建完成,进入S4。

[0045] 所述S3.2中的工艺语句分词模块的训练过程如下:

[0046] S3.2.1:利用分词包jieba对纯文本格式文件中的语句进行初步分词;

[0047] S3.2.2:人工对S3.2.1中得到的分词结果进行判断修改,标注“B”、“I”、“W”、“S”标签,其中“B”表示词的首汉字、“I”表示词的中间汉字、“W”表示词的末尾汉字、“S”表示词仅有一个汉字;

[0048] S3.2.3:将S3.2.2中得到的带有分词标签的文本存储到本地硬盘中作为工艺语句分词语料库;

[0049] S3.2.4:训练工艺语句分词模块,将S3.2.3得到的工艺语句分词语料库导入到工艺语句分词模块,然后读入分词模块程序进行迭代运行,程序运行完成后工艺语句分词模块训练完成。

[0050] 所述S3.3中工艺词语分类模块的训练过程如下:

[0051] S3.3.1:人工对上面处理好的文本中的词语标注“零件”、“设备”、“装备”、“特征”、“关系”、“工序”、“工步”、“属性”八个类别,分别用“p”、“m”、“e”、“f”、“r”、“d”、“s”、“a”表示,无类别词用“x”表示,上述“零件”、“设备”、“装备”、“特征”、“关系”、“工序”、“工步”七个类别与S2.2中的六类数据相对应,“属性”与S2.3中的类的属性相对应;

[0052] 加工对象数据对应“零件”;加工设备数据对应“设备”;加工方法数据对应“工序”、“工步”,零件加工包含多个工序,每个工序包含多个工步;加工装备数据对应“装备”;加工特征数据对应“特征”;语义关系数据对应“关系”;“属性”包含数值和名称;

[0053] S3.3.2:将S3.3.1中得到的带有分类标签的文本存储到本地硬盘中作为工艺词语分类语料库;

[0054] S3.3.3:训练工艺词语分类模块,将S3.3.2得到的工艺词语分类语料库导入到工艺词语分类模块,然后读入分类模块程序进行迭代运行,程序运行完成后工艺词语分类模块训练完成。

[0055] S4:可视化展示,把模式层和数据层结合起来构建知识图谱,把抽取到的三元组输入到Neo4j图数据库中,存储知识图谱并可视化展示。

[0056] 与现有技术相比,本发明的有益效果为:

[0057] 1、本发明一种数控加工领域工艺数据的知识图谱构建方法提供了一种数控加工领域工艺数据的管理、整合、存储方法,相较于传统的关系型数据库存储更具直观。融合了加工过程的工艺数据,以图的形式整合数据,可以利用图的算法来实现推理过程,获得工艺数据的更大更深层次的价值,发挥数据的作用。按照本方法构建的知识图谱可以实现更加智能的功能,存储大量的知识图谱,把知识图谱结合起来,存储成工艺数据知识图谱库,通过挖掘数据和推理,实现工艺路线生成,优化工艺过程。

[0058] 2、本发明一种数控加工领域工艺数据的知识图谱构建方法中从模式层,对数据进

行了分类将立体的机加工,抽象为二维的知识图谱,在构建模式层时对数控加工领域的工艺数据进行了整理归纳,把繁多复杂的数据归结为六大类数据,抽象为模式层中本体模型的六大类实体,对工艺数据的合理分类有利于从宏观上清晰明了的掌握工艺数据的类别构成,以自顶到底的角度理解工艺数据。

[0059] 3、本发明一种数控加工领域工艺数据的知识图谱构建方法中的数据层构建是在模式层的约束下完成的,提取的冗余信息会变少,提取到的信息会更加符合原始工艺数据中的信息;由于目前的工艺数据大量存在于非结构化的文本当中,这就对以后工艺数据的挖掘和智能化生成工艺路线带来了阻碍,所以要把这些工艺数据用一个更加规则的方式表达出来。本发明把存在于文本当中有效的工艺数据挖掘出来,转换为语义关系数据,利用语义关系数据构建知识图谱并存储在图数据库中。

附图说明

[0060] 图1是本发明知识图谱构建方法整体流程图。

[0061] 图2是本发明模式层的构建流程图。

[0062] 图3是本发明模式层的示意图。

[0063] 图4是本发明数据层的构建流程图。

[0064] 图5是本发明训练基于Bi-LSTM-CRF的工艺语句分词模型的具体流程图。

[0065] 图6是本发明训练基于word2vec的工艺词语分类模型的具体流程图。

[0066] 图7是本发明语义关系构建的具体流程图。

[0067] 图8是本发明以传动轴为示例的数控加工工艺数据知识图谱。

具体实施方式

[0068] 以下结合附图说明和具体实施方式对本发明作进一步详细的说明。

[0069] 参见图1至图7,一种数控加工领域工艺数据的知识图谱构建方法,包括以下步骤:

[0070] S1:建立原始资料数据库,查找数控加工领域相关的数据资料,并建立原始资料数据库;

[0071] S2:构建知识图谱模式层,人工抽取原始资料数据库中的工艺知识,结合数控加工工艺领域的专家知识建立本体模型,构建知识图谱模式层;

[0072] S3:构建知识图谱数据层、提取三元组,利用数据抽取模型抽取非结构化数据中的实体、属性、关系,构建知识图谱数据层;由于原始资料数据库中的数据均属于包含大量信息的非结构化工艺数据,要完成知识图谱的自动建立,则需要将包含大量信息的非结构化工艺数据的信息进行抽取,将非结构化数据转化为语义关系数据;故建立数据抽取模型自动抽取非结构化数据中的实体、属性、关系,构建知识图谱数据层;所述数据抽取模型包括工艺语句分词模块、工艺词语分类模块和语义关系构建模块;

[0073] S4:可视化展示,模式层和数据层结合起来构建知识图谱,存储知识图谱并进行可视化展示。

[0074] 所述S1:建立原始资料数据库中的数控加工领域相关的数据资料包括:零件的图纸数据、规范性手册、加工过程工艺卡片、数控机床和加工中心的数据、夹具量具定位装备的数据、工艺文件、工序卡片、专业书籍、工艺手册;将查找到的数控加工领域相关的数据资

料,存储到本地硬盘中作为原始资料数据库。

[0075] 所述S2:构建知识图谱模式层,首先人工抽取原始资料数据库中的工艺知识,结合数控加工工艺领域的专家知识建立本体模型,构建知识图谱模式层;

[0076] S2.1:分析抽取出来的工艺数据,结合专家知识对数据进行归纳;要建立本体模型并构建知识图谱模式层;知识图谱模式层通过本体模型方式来构建,本体模型通过本体语言OWL来描述;针对原始资料数据中的非结构化数据通过人工提取工艺知识,然后结合专家知识建立本体模型构建知识图谱模式层;本体模型是模式层的模型,本体语言是描述模式层的语言;

[0077] S2.2:数控加工设备的加工对象是零部件,分析零部件的CAD模型和数控加工工艺文档包含的零部件信息及加工制造信息,可以将其归纳为六类数据:加工对象数据、加工设备数据、加工方法数据、加工装备数据、加工特征数据以及语义关系数据;

[0078] 工艺数据可以总结为:加工对象数据、加工设备数据、加工方法数据、加工装备数据、加工特征数据和语义关系数据;

[0079] 加工对象数据是整个工艺数据的基础,描述了加工对象的编号、名称等文本信息,还描述了包括机械模型的材料、尺寸、要求的信息;加工设备数据主要指加工零件需要的数控机床或者加工中心,例如数控铣床、数控车床;加工方法数据是指加工零件的方法和操作,材料从毛坯到零件要经过的步骤,具体指加工工序和工步;加工装备数据指的是工艺装备,包括定位用到的定位装置、夹紧装置,加工过程中的量具辅具,如机床上的刀具;加工特征数据是指机床所完成的工序中具体的加工对象及要求,如面、孔、螺纹,加工要求包括精度要求和材料处理的要求;语义关系数据是指前面几种数据之间的关系,表示不同类别的数据之间的联系;

[0080] S2.3:为了保持模式层的准确性和完备性,把以上的六种数据抽象为六大类,其中加工对象数据、加工设备数据、加工方法数据、加工装备数据、加工特征数据以及语义关系数据依次对应:零件类、设备类、工艺类、装备类、特征类、关系类;工艺类继承有工序类和工步类,装备类继承有刀具类、夹具类、量具类、辅具类,特征类继承有形状特征类、精度特征类、技术特征类、管理特征类、材料特征类,其中,精度特征类继承有尺寸精度类、形状精度类、位置精度类、表面粗糙度类,关系类继承有contain类、order类;

[0081] 类的属性包括类属性和实例属性,类属性为各类及其子类共享,所有实例共享对应的类属性,实例属性仅为各个实例所有;零件类具有名称类属性,设备类具有名称类属性,工艺类具有名称类属性,装备类具有名称类属性,特征类具有名称类属性,关系类具有名称类属性,其他实例属性根据不同实体添加;

[0082] S2.4:知识图谱的模式层描述一些实体类及之间的关系,模式层根据S 2.3中抽象的实体类建立,通过表达式来描述,把抽象出来的实体类和关系建立本体模型,利用protégé软件建立和展示,模式层的知识通过本体模型来表示,本体模型可以用owl语言来描述;

[0083] 知识图谱的模式层的本体模型形式化表达为:

[0084] $KGPattern = \{Entity \cup Relation\}$

[0085] 其中:

[0086] $Entity = \{P \cup M \cup O \cup E \cup F\}$

[0087] $Relation = \{R\}$

[0088] $R = \{(P_i, contain, F_j) \cup (O_i, contain, M_j) \cup (O_i, contain, O_j) \cup (O_i, contain, E_j) \cup (O_i, contain, F_j) \cup (O_i, order, O_j) \cup \}$

[0089] 以上式子中:KGPattern指知识图谱模式层的形式化表达模型,Entity指模式层描述的实体集合,Relation指模式层描述的关系集合;P表示零件类,M表示设备类,O表示工艺类,E表示装备类,F表示特征类,R表示关系类。语义关系R包括contain、order,contain关系定义为零件类与特征类之间的包含关系,工艺类与设备类、装备类、特征类的包含关系,以及工艺类之间的包含关系,关系表示及描述为:($P_i, contain, F_j$)表示零件 P_i 包含特征 F_j , ($O_i, contain, M_j$)表示工艺 O_i 包含设备 M_j , ($O_i, contain, O_j$)表示工艺 O_i 包含工艺 O_j , ($O_i, contain, E_j$)表示工艺 O_i 包含装备 E_j , ($O_i, contain, F_j$)表示工艺 O_i 包含特征 F_j ;order关系定义为工艺类中工序之间或工步之间的顺序关系,关系表示及描述为:($O_i, order, O_j$)表示工序(工步) O_i 在前,工序(工步) O_j 在后;如表1所示;

[0090] 表2语义关系表示及描述

[0091]	语义关系	关系类表示	关系类描述
[0092]	<i>contain</i>	$(P_i, contain, F_j)$	零件 P_i 包含特征 F_j
		$(O_i, contain, M_j)$	工艺 O_i 包含设备 M_j
		$(O_i, contain, O_j)$	工艺 O_i 包含工艺 O_j
		$(O_i, contain, E_j)$	工艺 O_i 包含装备 E_j
		$(O_i, contain, F_j)$	工艺 O_i 包含特征 F_j
	<i>order</i>	$(O_i, order, O_j)$	工序 O_i 在前, 工序 O_j 在后或工步 O_i 在前, 工步 O_j 在后

[0093] 此时,知识图谱模式层构建完成,进入下一步S3。

[0094] 所述S3:构建知识图谱数据层、提取三元组,利用数据抽取模型抽取非结构化数据中的实体、属性、关系,构建知识图谱数据层;

[0095] 由于原始资料数据库中的数据均属于包含大量信息的非结构化工艺数据,要完成知识图谱的自动建立,则需要将包含大量信息的非结构化工艺数据的信息进行抽取,将非结构化数据转化为语义关系数据;故建立数据抽取模型自动抽取非结构化数据中的实体、属性、关系,构建知识图谱数据层;所述数据抽取模型包括工艺语句分词模块、工艺词语分类模块和语义关系构建模块;具体运行流程如下:

[0096] S3.1:把待处理的非结构化数据转化为纯文本格式文件,通过编程读入数据,把数据保存为txt格式文件;

[0097] S3.2:利用工艺语句分词模块分词,所述工艺语句分词模块基于Bi-LSTM-CRF算法实现,利用的工艺语句分词模块分词,把S3.1中得到的纯文本数据输入到训练好的工艺语句分词模块中,最终输出带有分词标签“B”、“I”、“W”、“S”的文本文件,此时文本分词完成:

[0098] S3.3:将分词完成的文本内的词语进行分类形成词向量,工艺词语分类模块基于word2vec算法实现,将S3.2得到的带有分词标签的文本文件中的分词作为待分类词语,计算待分类词语与加工词语分类语料库的词语的余弦距离,将距离最短的样本的类别赋予待分类词语相同的类别,最终输出带有词语类别标签的文本文件,此时词语分类完成;

[0099] S3.4:语义关系构建,将S3.3中得到的带有词语类别标签的文本文件,根据模式层定义的实体类别来识别词语,根据模式层定义的实体之间的关系来识别词语之间的关系,并以三元组方式存储,根据词语属性标签来识别实体和属性之间的关系,也以三元组方式保存,保存的三元组输出为<实体,关系,实体>,<实体,属性,属性值>三元组。

[0100] 所述S3.4:语义关系构建,将S3.3中得到的带有词语类别标签的文本剔除无类别的占位词,得到有类别的词语文本集合,按顺序依次处理类别的词语文本集合内的词语得到三元组输出,直到所有词语处理完成;所述处理过程如下:

[0101] S3.4.1:如前者词语标签为“p”,且下一个词语标签为“f”,则添加三元组<前者词语,contain,后者词语>,然后进入下一词语的处理;如果不是则进入S3.4.2;

[0102] S3.4.2:如前者词语标签为“p、m、e、f、d、s”中的任意一个,且下一个词语标签为“a”,则添加三元组<前者词语,ofattribute,后者词语>,然后进入下一词语的处理;如果不是则进入S3.4.3;

[0103] S3.4.3:如前者词语标签为“d”或“s”,且下一个词语标签为“m”、“e”或“f”,则添加三元组<前者词语,contain,后者词语>,然后进入下一词语的处理;如果不是则进入S3.4.4;

[0104] S3.4.4:如前者词语标签为“d”,且下一个词语标签为“s”,则添加三元组<前者词语,contain,后者词语>,然后进入下一词语的处理;如果不是则进入S3.4.4;

[0105] S3.4.5:如前者词语标签和下一个词语标签均为“d”,或前者词语标签和下一个词语标签均为“s”,则添加三元组<前者词语,order,后者词语>,然后进入下一词语的处理;如果不是则直接进入下一词语的处理;

[0106] 所有词语处理完成后将其以csv文件的格式保存,此时语义关系构建完成,进入S4。

[0107] 所述S3.2中的工艺语句分词模块的训练过程如下:

[0108] S3.2.1:利用分词包jieba对纯文本格式文件中的语句进行初步分词;

[0109] S3.2.2:人工对S3.2.1中得到的分词结果进行判断修改,标注“B”、“I”、“W”、“S”标签,其中“B”表示词的首汉字、“I”表示词的中间汉字、“W”表示词的末尾汉字、“S”表示词仅有一个汉字;

[0110] S3.2.3:将S3.2.2中得到的带有分词标签的文本存储到本地硬盘中作为工艺语句分词语料库;

[0111] S3.2.4:训练工艺语句分词模块,将S3.2.3得到的工艺语句分词语料库导入到工艺语句分词模块,然后读入分词模块程序进行迭代运行,程序运行完成后工艺语句分词模块训练完成。

[0112] 所述S3.3中工艺词语分类模块的训练过程如下:

[0113] S3.3.1:人工对上面处理好的文本中的词语标注“零件”、“设备”、“装备”、“特征”、“关系”、“工序”、“工步”、“属性”八个类别,分别用“p”、“m”、“e”、“f”、“r”、“d”、“s”、“a”表示,无类别词用“x”表示,上述“零件”、“设备”、“装备”、“特征”、“关系”、“工序”、“工步”七个类别与S2.2中的六类数据相对应,“属性”与S2.3中的类的属性相对应;

[0114] 加工对象数据对应“零件”;加工设备数据对应“设备”;加工方法数据对应“工序”、“工步”,零件加工包含多个工序,每个工序包含多个工步;加工装备数据对应“装备”;加工特征数据对应“特征”;语义关系数据对应“关系”;“属性”包含数值和名称;

[0115] S3.3.2:将S3.3.1中得到的带有分类标签的文本存储到本地硬盘中作为工艺词语分类语料库;

[0116] S3.3.3:训练工艺词语分类模块,将S3.3.2得到的工艺词语分类语料库导入到工艺词语分类模块,然后读入分类模块程序进行迭代运行,程序运行完成后工艺词语分类模块训练完成。

[0117] S4:可视化展示,把模式层和数据层结合起来构建知识图谱,把抽取到的三元组输入到Neo4j图数据库中,存储知识图谱并可视化展示。

[0118] 本发明的原理说明如下:

[0119] 从网站下载或企业内部获得上面所要求的资料,存储到本地硬盘中作为原始资料数据库,以待后续使用。

[0120] 知识图谱由模式层和数据层构成。模式层是经过归纳提炼的知识,构建模式层的模型有很多,本体模型是其中一种,本体模型的描述语言有好几种,OWL语言也仅是其中一种。数据层是实例化的模式层。相当于模式层中包含的仅仅是类名,比如电视剧类,而数据层就包含具体是什么,比如西游记。

[0121] 另外添加属性这一类别,属性并不属于模式层中的实体,但根据S2.3的最后一段可知,属性包含数值和名称。本专利并不把属性作为实体。比如设备是实体,那么设备名字属于设备这个实体的属性,并不把设备名字归为模式层中的某个实体。

[0122] S3.2.3中的分词语料库是为了训练分词模型的,里面的标签是每个字在词语中的位置,说的是文本的那几个字在一起可能是一个词语,但是并不知道这个词语是什么类别。本步骤是为了训练分类模型的,本步骤的语料库中的文本内容是按词语断开的,说的是词语可能是本专利规定的哪些类别。比如,“用插齿机加工齿轮”这句话,在S3.2.3的语料库的文本中是这样的:“用插齿机加工齿轮,S BIW BW BW”,而这句话在本步骤的语料库中的文本中是这样的:“用插齿机加工齿轮,x m x p”。两个语料库中的文本都是一句话为一行。

[0123] S4:可视化展示,把模式层和数据层结合起来构建知识图谱,此处的结合意指模式层对于数据层的约束,因为数据层中的词语类别都是模式层中规定的,词语之间的关系也是模式层规定的,所以在前文步骤词语分类和语义关系构建的时候就是结合了。

[0124] Bi-LSTM-CRF算法:(Bi-directional-Long Short Term Memory-Conditional Random Field)双向长短时记忆模型条件随机场算法。

[0125] 分词包jieba是软件python中的一个包名字,包名就是jieba。

[0126] word2vec:是一个双层的神经网络算法,用来把自然语言处理中的字词转换为稠密向量。在这个算法之前字词的表示用的是one-hot编码的方式,太过稀疏,会引发维度灾

难,而且忽略了字词之间的关系。word2vec把字词表示为稠密向量,利用向量距离反映字词之间的关系。word2vec算法在训练完成时,可以把输入的词映射为向量。

[0127] 实施例1:

[0128] 一种数控加工领域工艺数据的知识图谱构建方法,包括以下步骤:

[0129] S1:建立原始资料数据库,查找数控加工领域相关的数据资料,并建立原始资料数据库;

[0130] S2:构建知识图谱模式层,人工抽取原始资料数据库中的工艺知识,结合数控加工工艺领域的专家知识建立本体模型,构建知识图谱模式层;

[0131] S3:构建知识图谱数据层、提取三元组,利用数据抽取模型抽取非结构化数据中的实体、属性、关系,构建知识图谱数据层;由于原始资料数据库中的数据均属于包含大量信息的非结构化工艺数据,要完成知识图谱的自动建立,则需要将包含大量信息的非结构化工艺数据的信息进行抽取,将非结构化数据转化为语义关系数据;故建立数据抽取模型自动抽取非结构化数据中的实体、属性、关系,构建知识图谱数据层;所述数据抽取模型包括工艺语句分词模块、工艺词语分类模块和语义关系构建模块;

[0132] S4:可视化展示,模式层和数据层结合起来构建知识图谱,存储知识图谱并进行可视化展示。

[0133] 实施例2:

[0134] 实施例2与实施例1基本相同,其不同之处在于:

[0135] 所述S1:建立原始资料数据库中的数控加工领域相关的数据资料包括:零件的图纸数据、规范性手册、加工过程工艺卡片、数控机床和加工中心的数据、夹具量具定位装备的数据、工艺文件、工序卡片、专业书籍、工艺手册;将查找到的数控加工领域相关的数据资料,存储到本地硬盘中作为原始资料数据库。

[0136] 所述S2:构建知识图谱模式层,首先人工抽取原始资料数据库中的工艺知识,结合数控加工工艺领域的专家知识建立本体模型,构建知识图谱模式层;

[0137] S2.1:分析抽取出来的工艺数据,结合专家知识对数据进行归纳;要建立本体模型并构建知识图谱模式层;知识图谱模式层通过本体模型方式来构建,本体模型通过本体语言OWL来描述;针对原始资料数据中的非结构化数据通过人工提取工艺知识,然后结合专家知识建立本体模型构建知识图谱模式层;本体模型是模式层的模型,本体语言是描述模式层的语言;

[0138] S2.2:数控加工设备的加工对象是零部件,分析零部件的CAD模型和数控加工工艺文档包含的零部件信息及加工制造信息,可以将其归纳为六类数据:加工对象数据、加工设备数据、加工方法数据、加工装备数据、加工特征数据以及语义关系数据;

[0139] 工艺数据可以总结为:加工对象数据、加工设备数据、加工方法数据、加工装备数据、加工特征数据和语义关系数据;

[0140] 加工对象数据是整个工艺数据的基础,描述了加工对象的编号、名称等文本信息,还描述了包括机械模型的材料、尺寸、要求的信息;加工设备数据主要指加工零件需要的数控机床或者加工中心,例如数控铣床、数控车床;加工方法数据是指加工零件的方法和操作,材料从毛坯到零件要经过的步骤,具体指加工工序和工步;加工装备数据指的是工艺装备,包括定位用到的定位装置、夹紧装置,加工过程中的量具辅具,如机床上的刀具;加工特

征数据是指机床所完成的工序中具体的加工对象及要求,如面、孔、螺纹,加工要求包括精度要求和材料处理的要求;语义关系数据是指前面几种数据之间的关系,表示不同类别的数据之间的联系;

[0141] S2.3:为了保持模式层的准确性和完备性,把以上的六种数据抽象为六大类,其中加工对象数据、加工设备数据、加工方法数据、加工装备数据、加工特征数据以及语义关系数据依次对应:零件类、设备类、工艺类、装备类、特征类、关系类;工艺类继承有工序类和工步类,装备类继承有刀具类、夹具类、量具类、辅具类,特征类继承有形状特征类、精度特征类、技术特征类、管理特征类、材料特征类,其中,精度特征类继承有尺寸精度类、形状精度类、位置精度类、表面粗糙度类,关系类继承有contain类、order类;

[0142] 类的属性包括类属性和实例属性,类属性为各类及其子类共享,所有实例共享对应的类属性,实例属性仅为各个实例所有;零件类具有名称类属性,设备类具有名称类属性,工艺类具有名称类属性,装备类具有名称类属性,特征类具有名称类属性,关系类具有名称类属性,其他实例属性根据不同实体添加;

[0143] S2.4:知识图谱的模式层描述一些实体类及之间的关系,模式层根据S2.3中抽象的实体类建立,通过表达式来描述,把抽象出来的实体类和关系建立本体模型,利用protégé软件建立和展示,模式层的知识通过本体模型来表示,本体模型可以用owl语言来描述;

[0144] 知识图谱的模式层的本体模型形式化表达为:

[0145] $KGPattern = \{Entity \cup Relation\}$

[0146] 其中:

[0147] $Entity = \{P \cup M \cup O \cup E \cup F\}$

[0148] $Relation = \{R\}$

[0149] $R = \{(P_i, contain, F_j) \cup (O_i, contain, M_j) \cup (O_i, contain, O_j) \cup (O_i, contain, E_j) \cup (O_i, contain, F_j) \cup (O_i, order, O_j) \cup \}$

[0150] 以上式子中:KGPattern指知识图谱模式层的形式化表达模型,Entity指模式层描述的实体集合,Relation指模式层描述的关系集合;P表示零件类,M表示设备类,O表示工艺类,E表示装备类,F表示特征类,R表示关系类。语义关系R包括contain、order,contain关系定义为零件类与特征类之间的包含关系,工艺类与设备类、装备类、特征类的包含关系,以及工艺类之间的包含关系,关系表示及描述为:($P_i, contain, F_j$)表示零件 P_i 包含特征 F_j ,($O_i, contain, M_j$)表示工艺 O_i 包含设备 M_j ,($O_i, contain, O_j$)表示工艺 O_i 包含工艺 O_j ,($O_i, contain, E_j$)表示工艺 O_i 包含装备 E_j ,($O_i, contain, F_j$)表示工艺 O_i 包含特征 F_j ;order关系定义为工艺类中工序之间或工步之间的顺序关系,关系表示及描述为:($O_i, order, O_j$)表示工序(工步) O_i 在前,工序(工步) O_j 在后;如表1所示;

[0151] 表3语义关系表示及描述

语义关系	关系类表示	关系类描述
[0152] <i>contain</i>	$(P_i, contain, F_j)$	零件 P_i 包含特征 F_j
	$(O_i, contain, M_j)$	工艺 O_i 包含设备 M_j
	$(O_i, contain, O_j)$	工艺 O_i 包含工艺 O_j
	$(O_i, contain, E_j)$	工艺 O_i 包含装备 E_j
	$(O_i, contain, F_j)$	工艺 O_i 包含特征 F_j
<i>order</i>	$(O_i, order, O_j)$	工序 O_i 在前, 工序 O_j 在后或工步 O_i 在前, 工步 O_j 在后

[0153] 此时,知识图谱模式层构建完成,进入下一步S3。

[0154] 所述S3:构建知识图谱数据层、提取三元组,利用数据抽取模型抽取非结构化数据中的实体、属性、关系,构建知识图谱数据层;

[0155] 由于原始资料数据库中的数据均属于包含大量信息的非结构化工艺数据,要完成知识图谱的自动建立,则需要将包含大量信息的非结构化工艺数据的信息进行抽取,将非结构化数据转化为语义关系数据;故建立数据抽取模型自动抽取非结构化数据中的实体、属性、关系,构建知识图谱数据层;所述数据抽取模型包括工艺语句分词模块、工艺词语分类模块和语义关系构建模块;具体运行流程如下:

[0156] S3.1:把待处理的非结构化数据转化为纯文本格式文件,通过编程读入数据,把数据保存为txt格式文件;

[0157] S3.2:利用工艺语句分词模块分词,所述工艺语句分词模块基于Bi-LSTM-CRF算法实现,利用的工艺语句分词模块分词,把S3.1中得到的纯文本数据输入到训练好的工艺语句分词模块中,最终输出带有分词标签“B”、“I”、“W”、“S”的文本文件,此时文本分词完成:

[0158] S3.3:将分词完成的文本内的词语进行分类形成词向量,工艺词语分类模块基于word2vec算法实现,将S3.2得到的带有分词标签的文本文件中的分词作为待分类词语,计算待分类词语与加工词语分类语料库的词语的余弦距离,将距离最短的样本的类别赋予待分类词语相同的类别,最终输出带有词语类别标签的文本文件,此时词语分类完成;

[0159] S3.4:语义关系构建,将S3.3中得到的带有词语类别标签的文本文件,根据模式层定义的实体类别来识别词语,根据模式层定义的实体之间的关系来识别词语之间的关系,并以三元组方式存储,根据词语属性标签来识别实体和属性之间的关系,也以三元组方式保存,保存的三元组输出为<实体,关系,实体>,<实体,属性,属性值>三元组。

[0160] 所述S3.4:语义关系构建,将S3.3中得到的带有词语类别标签的文本剔除无类别的占位词,得到有类别的词语文本集合,按顺序依次处理类别的词语文本集合内的词语得

到三元组输出,直到所有词语处理完成;所述处理过程如下:

[0161] S3.4.1:如前者词语标签为“p”,且下一个词语标签为“f”,则添加三元组<前者词语,contain,后者词语>,然后进入下一词语的处理;如果不是则进入S3.4.2;

[0162] S3.4.2:如前者词语标签为“p、m、e、f、d、s”中的任意一个,且下一个词语标签为“a”,则添加三元组<前者词语,ofattribute,后者词语>,然后进入下一词语的处理;如果不是则进入S3.4.3;

[0163] S3.4.3:如前者词语标签为“d”或“s”,且下一个词语标签为“m”、“e”或“f”,则添加三元组<前者词语,contain,后者词语>,然后进入下一词语的处理;如果不是则进入S3.4.4;

[0164] S3.4.4:如前者词语标签为“d”,且下一个词语标签为“s”,则添加三元组<前者词语,contain,后者词语>,然后进入下一词语的处理;如果不是则进入S3.4.4;

[0165] S3.4.5:如前者词语标签和下一个词语标签均为“d”,或前者词语标签和下一个词语标签均为“s”,则添加三元组<前者词语,order,后者词语>,然后进入下一词语的处理;如果不是则直接进入下一词语的处理;

[0166] 所有词语处理完成后将其以csv文件的格式保存,此时语义关系构建完成,进入S4。

[0167] 实施例3:

[0168] 实施例3与实施例2基本相同,其不同之处在于:

[0169] 所述S3.2中的工艺语句分词模块的训练过程如下:

[0170] S3.2.1:利用分词包jieba对纯文本格式文件中的语句进行初步分词;

[0171] S3.2.2:人工对S3.2.1中得到的分词结果进行判断修改,标注“B”、“I”、“W”、“S”标签,其中“B”表示词的首汉字、“I”表示词的中间汉字、“W”表示词的末尾汉字、“S”表示词仅有一个汉字;

[0172] S3.2.3:将S3.2.2中得到的带有分词标签的文本存储到本地硬盘中作为工艺语句分词语料库;

[0173] S3.2.4:训练工艺语句分词模块,将S3.2.3得到的工艺语句分词语料库导入到工艺语句分词模块,然后读入分词模块程序进行迭代运行,程序运行完成后工艺语句分词模块训练完成。

[0174] 所述S3.3中工艺词语分类模块的训练过程如下:

[0175] S3.3.1:人工对上面处理好的文本中的词语标注“零件”、“设备”、“装备”、“特征”、“关系”、“工序”、“工步”、“属性”八个类别,分别用“p”、“m”、“e”、“f”、“r”、“d”、“s”、“a”表示,无类别词用“x”表示,上述“零件”、“设备”、“装备”、“特征”、“关系”、“工序”、“工步”七个类别与S2.2中的六类数据相对应,“属性”与S2.3中的类的属性相对应;

[0176] 加工对象数据对应“零件”;加工设备数据对应“设备”;加工方法数据对应“工序”、“工步”,零件加工包含多个工序,每个工序包含多个工步;加工装备数据对应“装备”;加工特征数据对应“特征”;语义关系数据对应“关系”;“属性”包含数值和名称;

[0177] S3.3.2:将S3.3.1中得到的带有分类标签的文本存储到本地硬盘中作为工艺词语分类语料库;

[0178] S3.3.3:训练工艺词语分类模块,将S3.3.2得到的工艺词语分类语料库导入到工

艺词语分类模块,然后读入分类模块程序进行迭代运行,程序运行完成后工艺词语分类模块训练完成。

[0179] S4:可视化展示,把模式层和数据层结合起来构建知识图谱,把抽取到的三元组输入到Neo4j图数据库中,存储知识图谱并可视化展示。

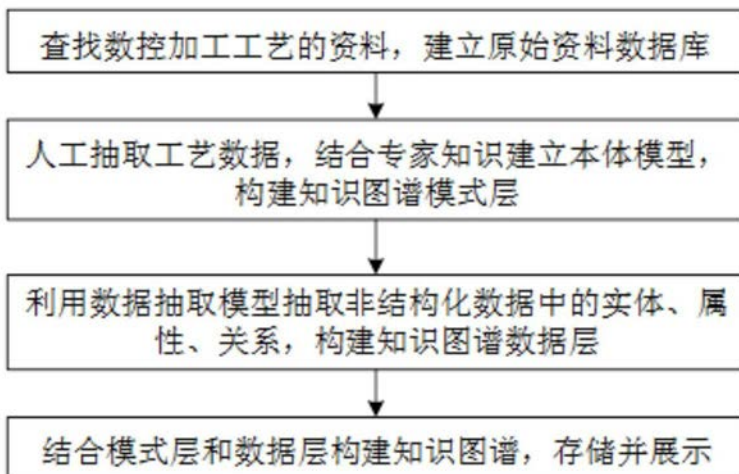


图1

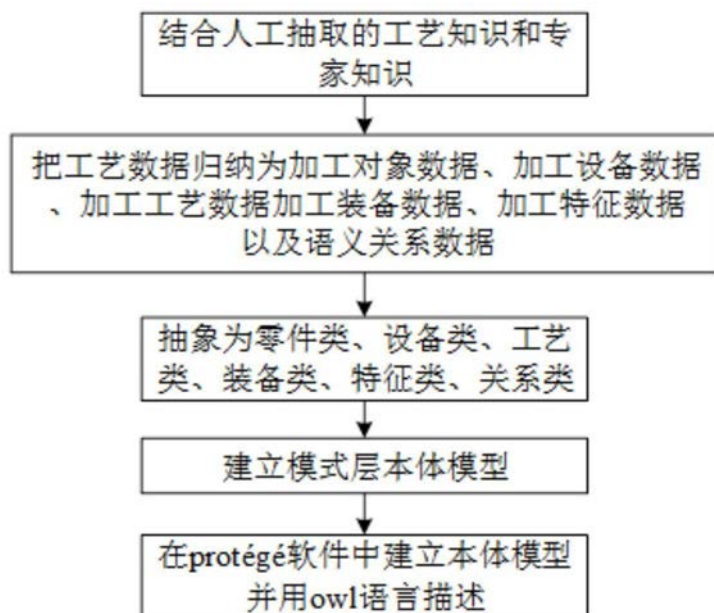


图2

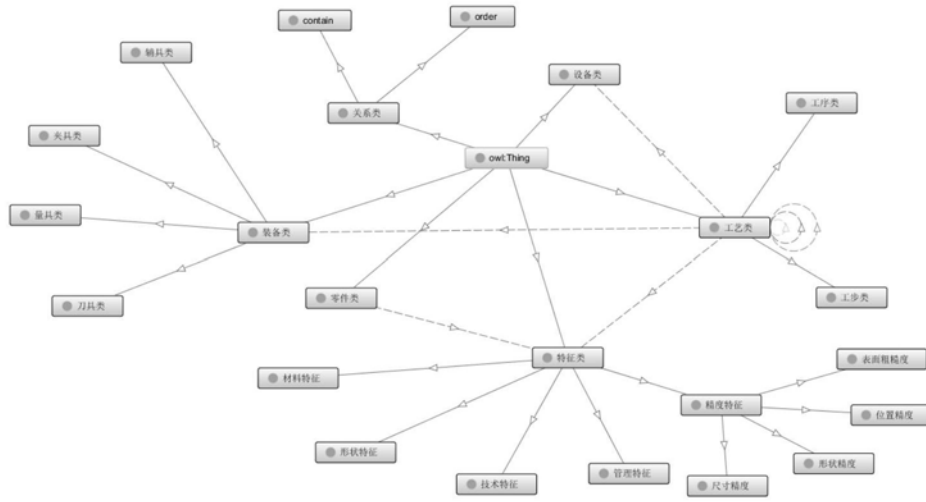


图3

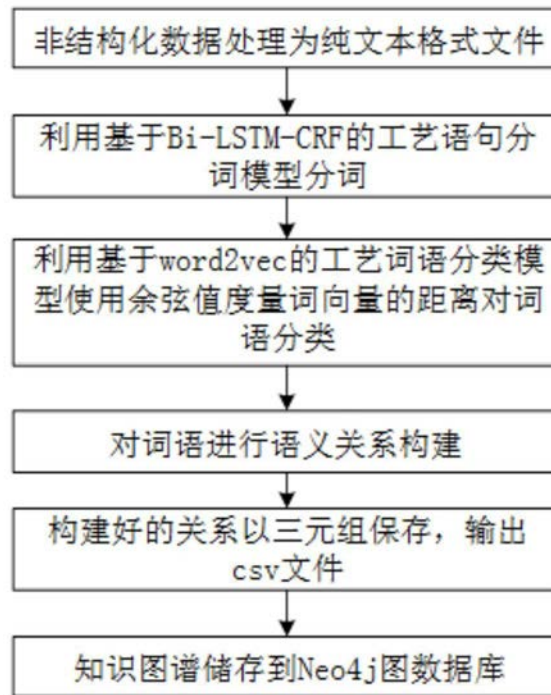


图4

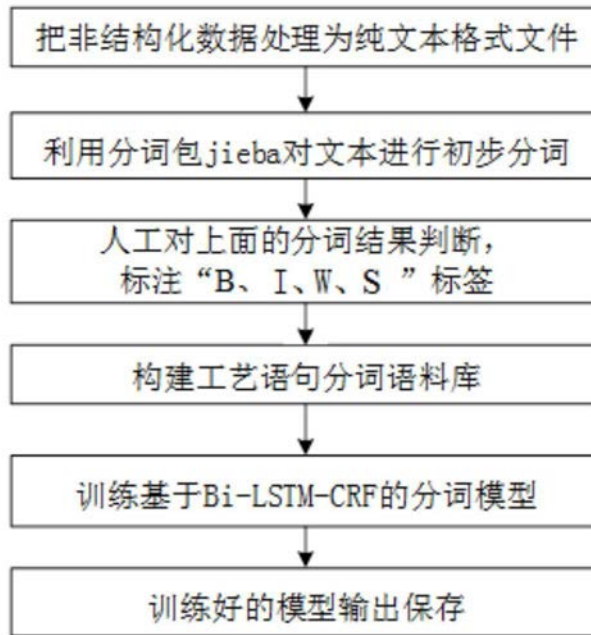


图5

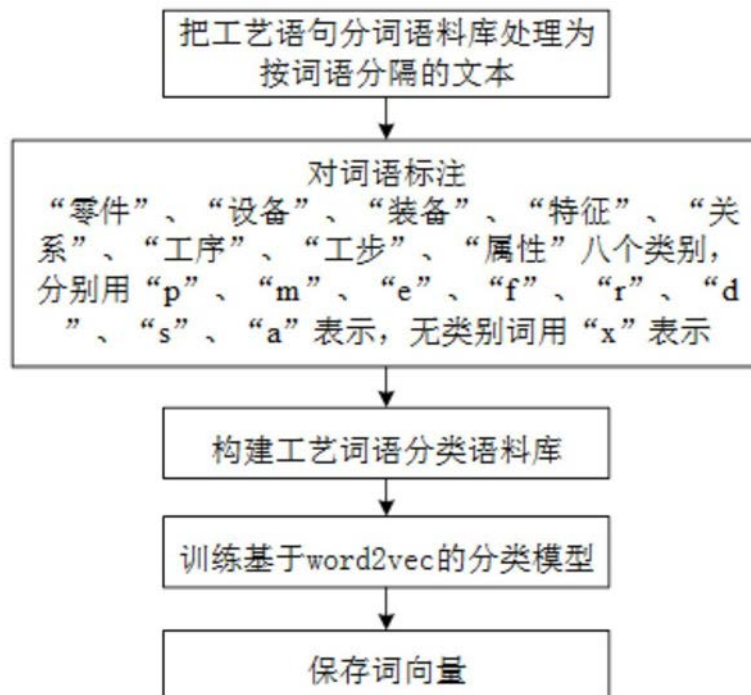


图6

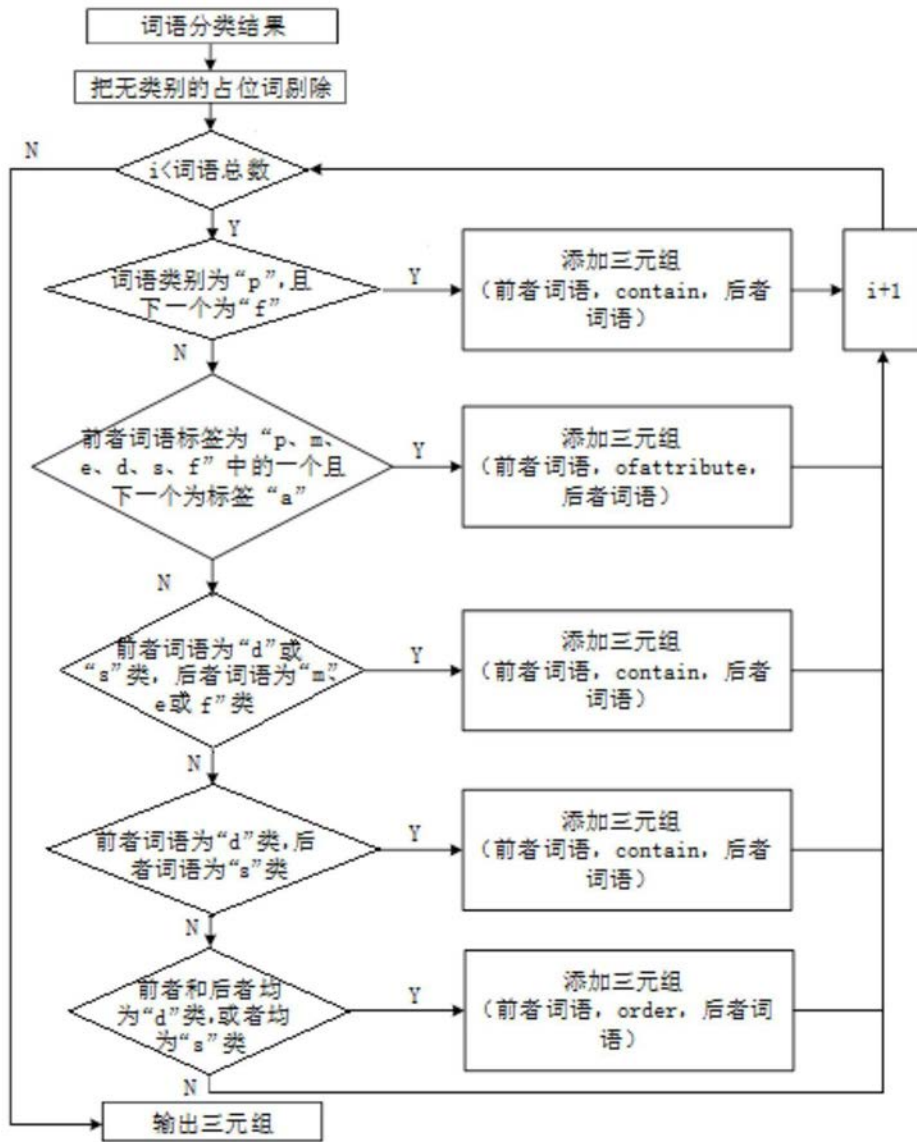


图7

