

doi:10.3772/j.issn.1000-0135.2015.004.007

基于语义分析的产品分类本体查询研究¹⁾

黄奇¹ 郑建国²

(1. 南京大学国家信息资源管理南京研究基地, 南京 210093;
2. 南京大学工程管理学院, 南京 210093)

摘要 本文在充分理解产品分类的重要意义及分析研究现状的基础上,剖析了产品分类的体系结构;针对产品分类固有的五大缺陷,提出了产品分类语义分析模型,并依据语义分析的结果构建了产品分类本体。为了验证产品分类本体推理的可判定性、概念的正确性和语义的完备性,建立了产品分类本体查询模型;按照真实查询的要求,该模型由本体解析、用户请求、查询处理、查询扩展、推理规则和查询结果反馈等模块组成;通过编程对该模型予以实现,建立了相应的实验系统;对实验系统进行运行,结果证明了需要验证的“三性”是成立的。

关键词 语义分析 本体构建 本体查询 GPC eCI@ss

Ontology Query for Product Classification Based on Semantic Analysis

Huang Qi¹ and Zheng Jianguo²

(1. State Information Resources Management Research Base, Nanjing University, Nanjing 210093;
2. School of Management Engineering, Nanjing University, Nanjing 210093)

Abstract Based on fully understanding the importance of product classification and analyzing the research status, we analyze the structure of the product classifications; Based on five inherent defects of product classification, we propose a model of product classification using semantic analysis method. According to the results of the semantic analysis, we construct a product classification ontology. In order to verify reasoning decidability of completeness, correctness and semantic concept of the product classification ontology, we construct a query model for product classification ontology. The model is composed of the requirements of real queries which are made up of the ontology, the user requests, query processing, query expansion, reasoning rules and query results feedback modules. The model is realized by programming and corresponding experimental system is established. We ran the experimental system, and verified the three characteristics in our hypothesis.

Keywords semantic analysis, ontology construction, ontology query, GPC, eCI@ss

1 引言

产品分类是现代物流的基础之一,目前正在蓬勃发展的电子商务以及新兴的物联网都离不开产品

分类。建立完善、健全的物流网络体系的必要条件是:有系统性、一致性、实时共享的产品信息支撑。产品信息的复杂性在于产品种类繁多,涉及各种行业,不同行业需要不同的产品属性,即使是同行业中的产品,由于类型不同,其属性描述也可能不同,而

收稿日期:2014年11月23日。

作者简介:黄奇,男,1961年生,双学士,教授,主要研究方向:网络信息资源的语义化、电子商务。E-mail:hq@nju.edu.cn; 郑建国,男,1990年生,在读研究生,主要研究方向:网络信息资源的语义化、电子商务。E-mail:jianguozheng@126.com。

1) 本文系国家自然科学基金重点项目“基于知识组织的产品分类本体研究”(项目编号:13ATQ005)和教育部人文社会科学研究规划基金项目“产品分类的语义分析及其形式化研究”(项目编号:11YJA870014)研究成果之一。

现在连最基本的产品分类也不统一。例如胶鞋在生产领域属橡胶制品,与胶管、胶带分在一起,而在流通领域属鞋类,与皮鞋、布鞋分在一块儿,相差甚远。由于缺乏统一的产品分类标准及规范化的信息描述,人们犹如缺少一种交流语言,无法共享产品信息。

产品分类是进行语义分析的基础,选择合适的产品分类法是展开语义分析的关键,也是本体构建的基础。从本质上讲,语义分析就是对产品分类的进一步细化,这种细化主要体现在产品知识表达的广度和精度,语义分析作为连接产品分类和本体构建的纽带,在基于语义分析的本体构建中起着承上启下的作用。本体构建作为语义分析的最终目标,是语义分析的产物,语义分析的质量越高,构建的相应本体质量也就越高。因此要获得高质量的产品分类本体,就离不开语义分析。现有的产品分类本体不是用来取代现行的产品分类法,而是作为用户与产品分类之间的桥梁,使用户的查询通过产品本体映射到产品分类法中去;作为不同产品分类法之间的桥梁,使不同产品分类法在语义匹配的基础上进行相互转换。因此,一个合格的产品分类本体至少做到:推理的可判定性、概念的正确性和语义的完备性;且通过查询才能检验该本体是否满足这三个要求。

在 2004 年 IEEE 国际电子商务技术会议上,与会学者就提出了语义化产品分类的问题,Dongkyu Kim 等以《全球标准产品与服务分类》(UNSPSC)与 eCl@ss 为例,从形式语义、属性描述、树形结构与图形结构等建模的不同方面,分析了分类模式的种类,提出了语义分类模型,具体实现了类目管理系统,为语义分类体系建立打下基础,为本体概念的引入产品分类埋下了伏笔^[1]。但这种模型比较复杂,创建语义分类的效率很低。之后,Hyunja Lee 等将本体模型应用到电子商务领域,利用扩展实体联系和描述逻辑来构建电子目录本体,但并未建立起能够实现自动映射的工具,未在操作层面上实现本体的应用^[2]。至此,有关产品本体问题的研究逐步扩展开来。其中,最为著名的是 Martin Hepp^[3-5]对产品与服务本体描述问题的研究,其在 eCl@ss 分类体系的基础上描述产品的类别及属性,构建了一个产品本体,并指出了构建和使用本体的六大困难以及描述产品的本体应涵盖的六个方面。然而,Martin 构建的本体仅仅是用网络本体语言将 eCl@ss 分类系统书写了一遍,缺少对原有产品分类体系的再加工

与知识的提炼,没有对概念之间的语义关系做深入分析,对原产品分类体系中正确与错误的语义关系都予承认,并保留在新建的本体中。

近年来关于产品分类的研究热点主要集中在产品分类查询、产品分类体系的研究、产品分类语义化的研究、产品知识管理的研究和产品本体语义互操作的研究等^[6-14]。如 Rosati 等^[15]开发了基于本体的产品系统(GOPS),定义了 GOPS 的子系统 Lite-GOPS,使用轻量级的本体查询语言(EQL-Lite(UCQ))对 Lite-GOPS 的数据进行查询,并对子系统推理的可判定性和易用性进行了证明;Huang 等^[16]采用基于语义微分法设计产品分类,在电子商务产品分类设计中遵循顾客的情感需要;Beneventano 等^[17]提出了一套半自动创建语义分析的方法,该方法可在不同的产品分类体系之间实现语义映射,并利用 UNSPSC、eCl@ss 及 eBay 的部分分类目录进行了实际操作展示;Borsato^[18]通过构建产品本体,为产品生命周期管理与产品可持续发展两者之间建立了桥梁;Sorathia 等^[19]采用系统论方法来重点处理由于地区差异化而引发的本体互操作问题,并对针对现有本体开展了评估研究。目前国内外有关产品分类本体查询的文章鲜见,本文就是在上述研究的基础上,提出了一种基于语义分析模型的产品分类查询方法。

2 产品分类的结构

产品分类体系拓扑结构普遍采用了简单的树形结构,目前全球使用最广泛的产品分类法是 GPC(全球产品分类),而属性最丰富的分类法是 eCl@ss(先进分类系统)。本文主要介绍 GPC 和 eCl@ss 的体系结构。

GPC 和 eCl@ss 都采用四级类目结构。GPC 的 2013 标准共有 37 个大类、3812 个类目^[20];eCl@ss 6.2 版共有 27 个专业,包括 584 个主类,4982 个分类,27 200 个支类,14 371 个标准属性和 51 327 关键词^[21]。GPC 和 eCl@ss 每层用两位阿拉伯数字表示其类目编码,从第一层逐层分到第四层,类目编码由左往右连续排列,共有八位。GPC 的第 1 层表示产品的大类,为产品隶属的行业;第 2 层表示产品的中类,为产品隶属小行业;第 3 层表示产品的小类,为产品隶属的族;第 4 层表示产品的细类,具体的产品品种即基础产品类别。eCl@ss 的第 1 层为产品的专业范围;第 2 层为产品的主类;第 3 层为产品的

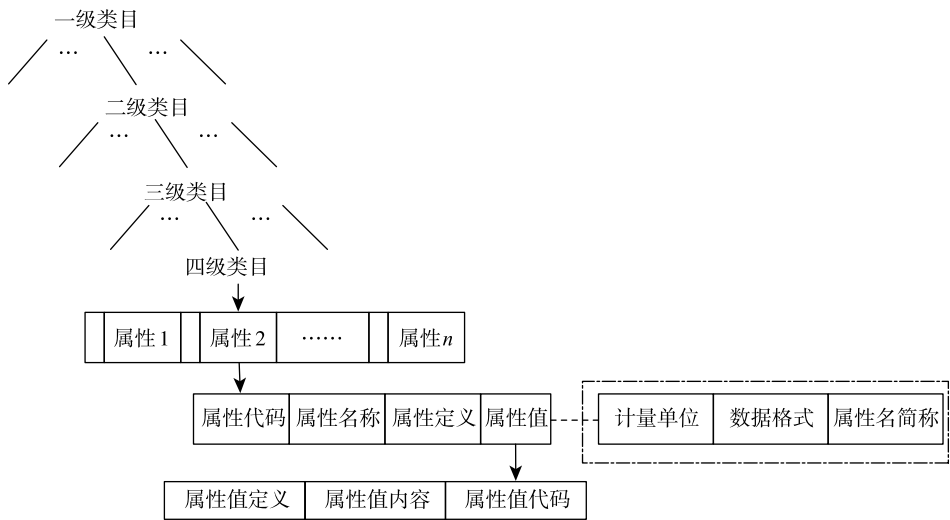


图 1 GPC 和 eCl@ss 产品分类体系结构

分类;第 4 层为产品的支类,具体的产品品种即基础产品类别。GPC 体系中第四层定义了属性;属性的描述字段有属性代码、属性名称、属性定义和属性值组成;属性值的描述字段有属性值代码、属性值定义和属性值内容组成。eCl@ss 的属性是专门定义的,分为基础属性(BSP)和专有属性(SSP),属性与 GPC 一样也定义在第四级类目上;每条属性的描述字段仅在 GPC 的基础上增加了属性名简称、数据格式和计量单位三个字段,其属性值的描述字段亦与 GPC 的相同。两者之间的结构如图 1 所示。图 1 中虚线框内容为 eCl@ss 所特有,其余为两者共有。

3 产品分类本体构建

3.1 产品分类本体构建模型

产品分类采用的是等级体系分类法,因此等级体系的系统性缺陷它也具有。归纳起来,产品分类主要存在如下五大问题^[22-24]:

- 类目线性排列与网状关系的矛盾。对于功能单一的产品,线性排列非常适用;而对于多功能、多用途的产品,线性排列的局限性是明显的;对于由通用部件组成的整机,线性排列会带来大量的重复或冗余。
- 语义表达能力有限。等级体系最能表达的是上下位类与同位类的关系。但上下位类中究竟是种属关系、整体与部分关系、整机与配件关系、概念与实例关系却无法确定。
- 概念的专指度受限。从等级分类体系原理上讲,概念的划分可以一直进行下去,使类目的末级

达到很高的专指度。但产品分类中受层级数的限制,末级类目划分往往较混乱。

- 使用过于专业。分类法中隐含的联系和限定贯穿于整个分类体系,对用户来说缺少易用性,不但有些类目本身的含义和类目间的逻辑联系对中小企业因雇不起这方面的专业人员而无法理解和掌握,其中有些技术甚至连专业人员也难以掌握。
- 周期长、更新慢。分类法主要以产品实体为分类对象,以逻辑划分为基础,兼顾产品属性。其产品范围划分相对稳定,体系严密而深细,更新周期比较长。

针对这些问题,本文设计了产品分类语义分析模型进行产品分类本体的构建,着重分析产品分类本体的属性、类和语义关系的处理。本文的产品分类本体构建模型如图 2 所示。

从图 2 可以看出该模型主要包括:概念划分、类处理、属性处理和语义细化四个模块。概念划分是进行类处理和属性处理的基础;类处理和属性处理是对概念特征的进一步分析;语义细化是链接属性处理和类处理的纽带,是对类与类、类与属性和属性与属性等关系的深度挖掘。

(1)概念划分:概念包含属性和类,概念划分就是通过分析概念特征(包括外延特征和内涵特征),将其划分为属性或类。处理的主要流程是获取的产品分类信息源,其包括网络信息资源,从中提取概念,补充概念,使产品概念集尽可能的完备,去除重复概念,得到经过整合的概念集。对整合的概念进行特征提取,将这些特征区分为内涵特征和外延特征。结合产品分类本体的现实语义需求,将其划分为属

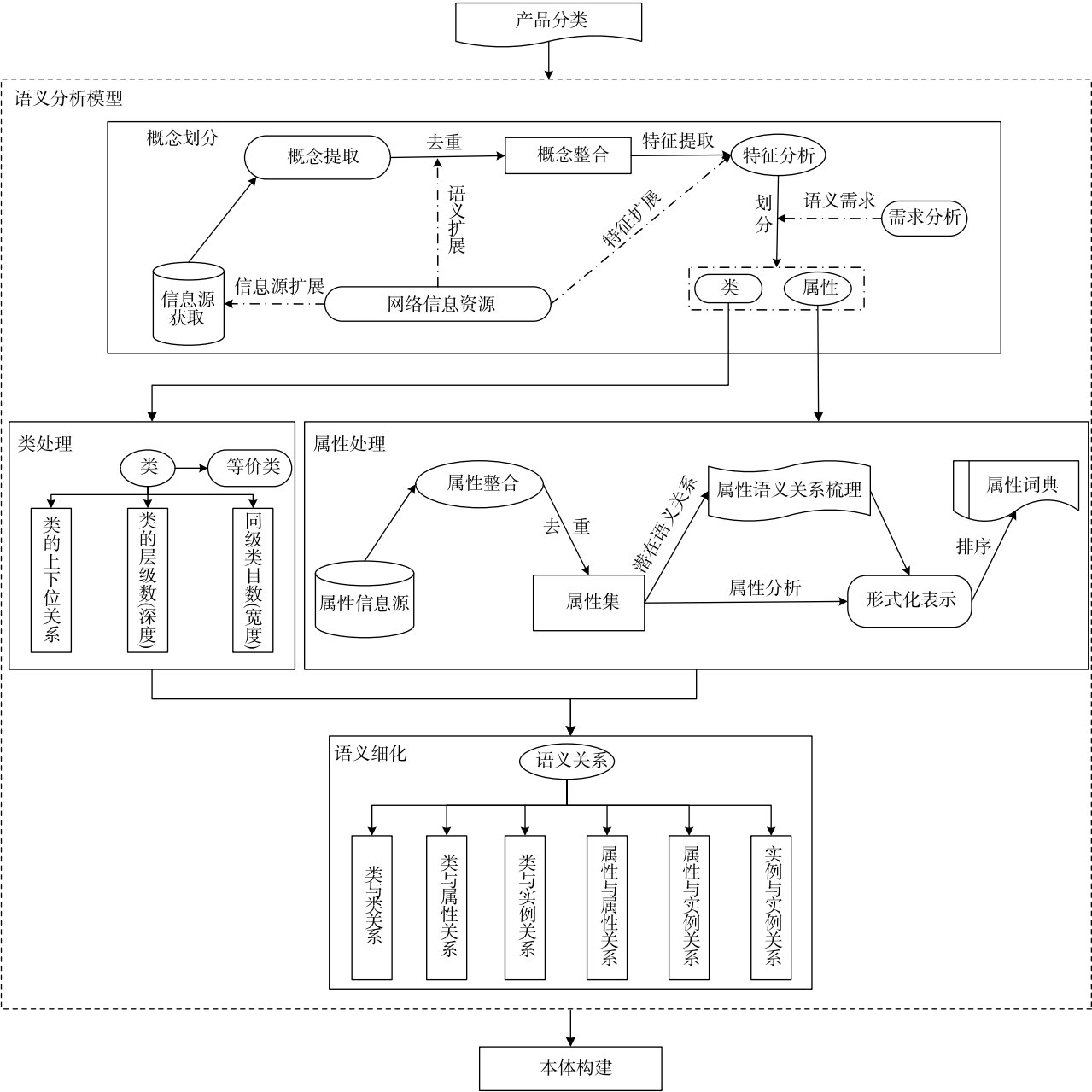


图2 产品分类本体构建模型图

性或类。

(2)类处理:指对获取的类进行细化和调整,是语义进一步丰富的过程。主要包括:确定类的上下位关系、确定类的层级数和同级类数目、确定类间关系和确定等价类。对于类的上下位关系采用自上而下的方法进行处理,对最上层的类按照其外延特征划分为下一级类,以此类推得到所有类的上下位关系。对特征丰富、语义需求高的类目要增加层级数和同级类数目;对特征欠缺、语义需求低的类目要减少层级数和同级类数目。对于类间关系,利用公理实现。对于等价类从产品设计角度出发,分析产品各组件的关联性,判断类间的等价关系。

(3)属性处理:指对属性集进行语义分析,构建属性词典。处理的主要流程是首先获取属性信息源,属性信息源主要来自于概念划分,然后对其进行去重,得到属性集。其次是对属性集进行分析,分析属性集中属性潜在的语义,进行属性的排序,构建属性词典。主要包括属性整合、属性语义关系梳理和形式化表达。属性整合指对概念划分得到的属性进行整合,去重得到属性集的过程;属性语义关系梳理指分析属性集各属性间的潜在语义关系,决定属性与属性间的内部层次结构;形式化表示指利用 OWL

语言描述属性。

(4) 语义细化:指的是考虑类、属性和实例相互之间的语义关系,通过语义分析对它们间的关系进行细化。按照语义细化的对象来分,语义关系可概括为类与类、类与属性、类与实例、属性与属性、属性与实例、实例与实例等。例如,上下位类中还可以进一步区分出种属关系(Kind-of)、部分与整体(Part-of)、整机与配件(Accessory-of)和物品与用料(Materials-of)等关系;并且类与属性、类与实例、属性与属性、属性与实例、实例与实例之间,也存在类似关系。再例如,通过不同部件类的组合(Make-up)可以构成最终产品(见例1)。

本体构建解决了产品分类的一些固有问题。通过类处理、语义细化解决了类目线性排列和网状关系的矛盾;通过语义细化解决了语义表达能力有限的问题;通过类处理解决了概念专指度受限的问题;通过概念划分为解决使用过于专业的问题奠定了基础,但问题真正的解决还有待查询系统的建立;周期长、更新慢的问题仅通过产品分类本体无法得到有效解决。此结论可以由图3验证。图3是GPC的Computing和Communication大类本体关系的全局示意图,其不仅有非常漂亮的树形结构(类目线性排列),而且存在复杂的网状关系;语义关系丰富(不同语义关系用不同颜色的线表示);使GPC原有的四级类目变成了五级类目,提高了概念的专指度。

3.2 产品分类本体构建的实现

本文选取了GPC的Computing和Communication大类和eCl@ss的第十九大类Information, communication and media technology进行产品分类本体构建。概念划分阶段共整理出598个GPC概念和2454个eCl@ss概念;其中598个GPC概念,均被划分为类目;而2454个eCl@ss概念,有1316个概念被划分为属性,其余的1138个概念均被划分为类。类处理和属性处理阶段,分别根据上述分类结果进行类和属性的梳理,GPC是采用现成的GDD属性词典;eCl@ss是对获取的1316个概念按照图2的模型分析后,对属性进行层次分类,构建属性词典。在语义细化阶段,构建了Part-of、Kind-of和Make-up等语义关系。

例1,类Smartphone(智能手机)由类Battery(电池)、CPU(中央处理器)、GPU(图形处理器)、RAM(内存)、Screen(显示屏)、Sensors(传感器)和Webcam(摄像头)等组成;类Desktop_Computer(台

式机)由类Motherboard(主板)、CPU(处理器)、RAM(内存)、Graphics(显卡)、Hard_disk(硬盘)、CD-ROM(光驱)、Cooling_Equipment(散热设备)、Mouse(鼠标)、Keyboard(键盘)、Webcam(摄像头)、Monitor(显示器)和Sound(音响)等组成。其中两者共有的部件是CPU、RAM、Screen和Webcam。为了减少冗余,将Smartphone(智能手机)和Desktop_Computer(台式机)中组成部件放到共用类Parts(部件类)中,利用类间关系Make-up(类组合关系),将Parts内的部件与智能手机和台式机相关联。这样从形式上仍为线性关系,而语义上则为网状关系,从而解决类目线性排列与网状关系矛盾。部分代码如下:

```
< owl:ObjectProperty rdf:ID = "make-up" >
    < rdfs:domain >
        < owl:Class >
            < owl:unionOf rdf:parseType = "
Collection" >
                < owl:Class rdf:about = "#Battery" / >
                < owl:Class rdf:about = "#CPU" / >
                < owl:Class rdf:about = "#GPU" / >
                < owl:Class rdf:about = "#RAM" / >
                < owl:Class rdf:about = "#Screen" / >
                < owl:Class rdf:about = "#Sensors" / >
                < owl:Class rdf:about = "#Webcam" / >
            < /owl:unionOf >
        < /owl:Class >
    < /rdfs:domain >
    < rdfs:range rdf:resource = "#Smartphone" / >
< /owl:ObjectProperty >
```

这段代码采用OWL-DL语言,定义了对对象属性Make-up,Make-up描述的是类与类之间潜在语义关系,其含义是类Battery(电池)、CPU(中央处理器)、GPU(图形处理器)、RAM(内存)、Screen(显示屏)、Sensors(传感器)和Webcam(摄像头)可以通过Make-up关系组成类Smartphone(智能手机)。

4 产品分类本体查询

产品分类本体查询就是要验证所构建的产品分类本体是否合格。一个合格的产品分类本体需要满足三个条件:一是推理的可判定性;二是概念的正确性;三是语义的完备性。

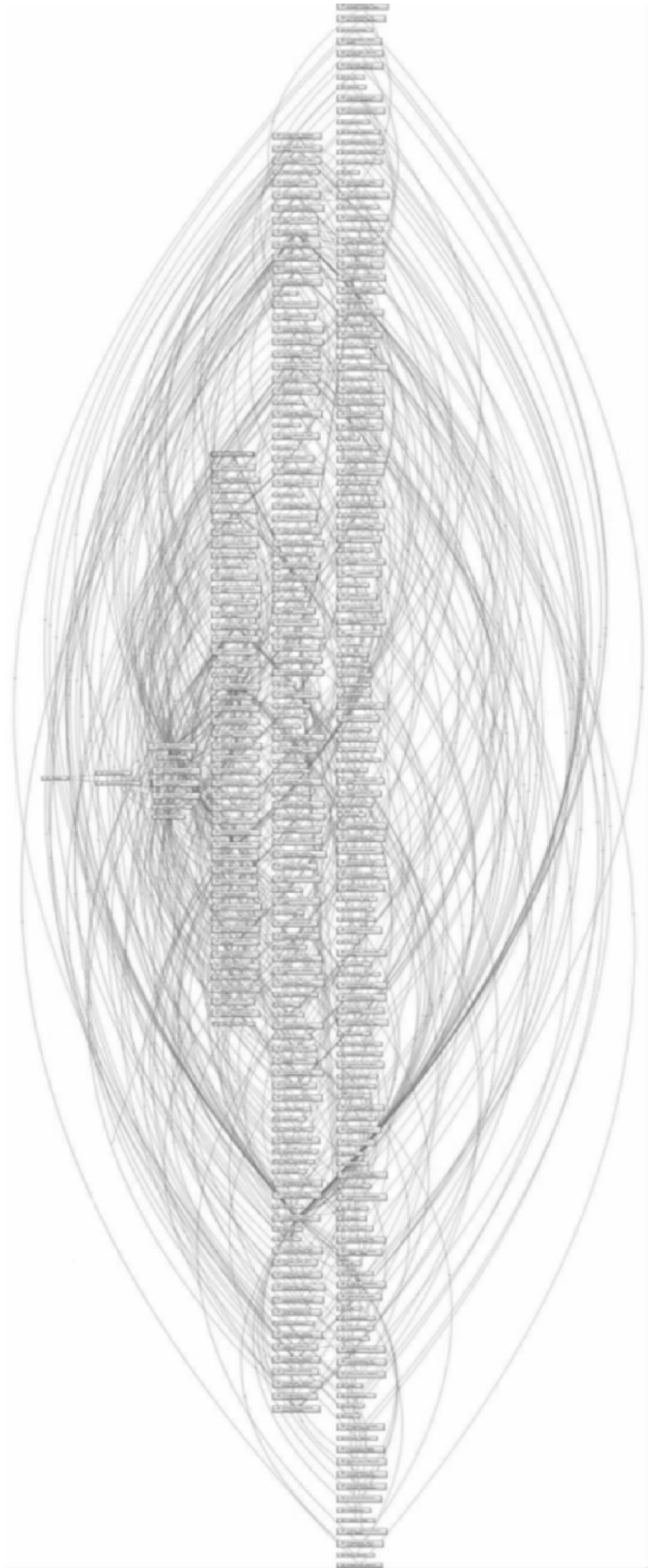


图3 GPC 的 Computing 和 Communication 大类本体关系的全局示意图

4.1 产品分类本体查询模型

传统的产品分类查询方法,主要是利用目录、索引和关键词进行简单的形式匹配,只能说明产品分类文档中有相同表示形式的内容,不能说明该表示形式的含义为何,也就是说不涉及任何语义。产品分类本体查询是从简单的形式匹配上升到语义匹配,语义匹配通过推理从本体知识库中获取相关的知识,实现语义层面的知识检索,同时还可以加入推理规则,对产品分类体系结构进行语义扩展。

推理的可判定性指的是查询模型要在有限的时间内给出查询结果;然而某些产品本体可能出现环状结构,若使用一般的查询模型则不能保证所有的本体可在有限的时间内得到查询结果,如 OWL Full 语言允许用户使用环状结构构建产品分类本体。

概念的正确性指的是经过查询返回的产品分类类目要正确;然而由于某些产品分类本体采用的是自动化方法进行构建,所有的概念都是采用基于概率方法自动抽取生成的,所构建的产品分类本体正确率得不到有效的保证,而产品分类对概念的正确性要求相当苛刻;并且即便是人工手动构建的本体也普遍存在大小写不区分、使用特殊符号和输入错误等问题,在类目层级较多的本体中,这种情况尤为明显,若仅仅使用传统的本体查询模型,用户若不熟知产品本体的构建规则,则很可能无法得到正确的结果。

语义的完备性有两层涵义:一是指查询概念返回的产品种类要齐备;二是要满足用户所有的查询需求。然而在现有的产品本体查询模型中,以针对概念的子类和父类的查询为主,没有进行查询的扩展,查询的结构是线性的而非网状的,如图 4 所示,用户只能查询某一概念与之直接相关的子类和父类,而不能查询概念的同位兄弟类、间接子类和间接父类等,限制了产品分类本体的应用。用户的需求是多种多样的,语义完备的本体应该能满足用户的多种查询请求:如对不同形式的同义词和一词多义等。

本文的产品分类本体查询模型就是为了满足上述三个条件。其基本原理是利用 Jean 推理机和 SparQL 语言对本体进行解析,然后将结果返回给用户。该模型主要包括本体解析模块、查询处理模块、用户请求模块、查询扩展模块、规则加入模块和查询结果反馈模块等。本体解析是查询处理的基础,用户请求是查询处理的依据,查询处理是查询扩展的前提。它们之间具体的构成关系如图 5 所示。

4.1.1 本体解析模块

本体解析模块指的是通过本体加载模型加载产品分类本体,读取和解析本体,转变为可存储的数据结构。本体解析是本体查询的基础,亦是保证本体推理可判定性基础。主要流程如图 6 所示。

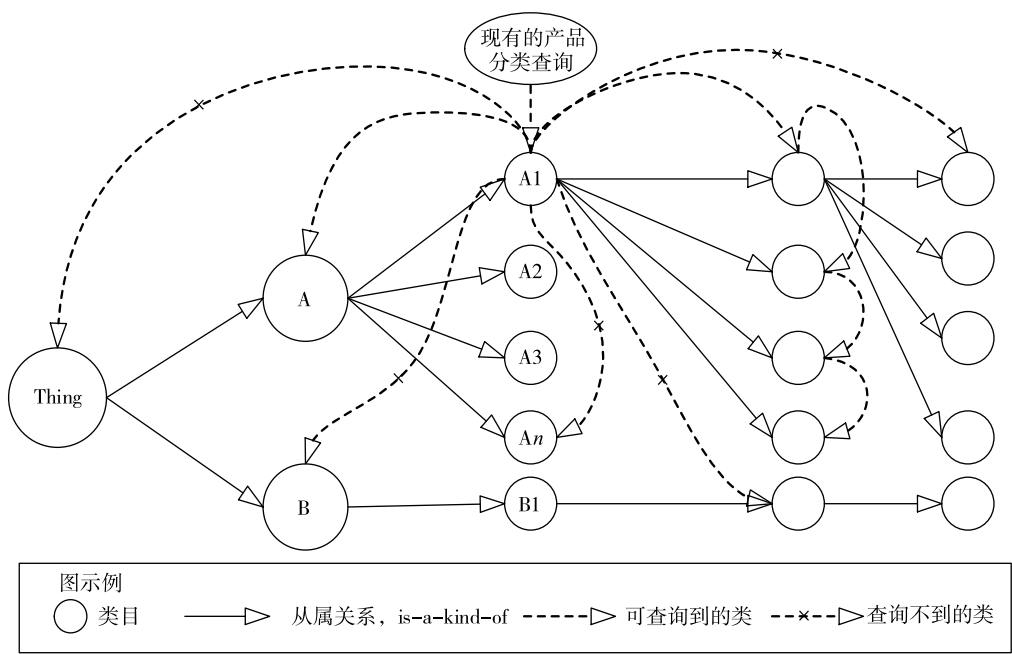


图 4 现有的产品分类查询主要模式

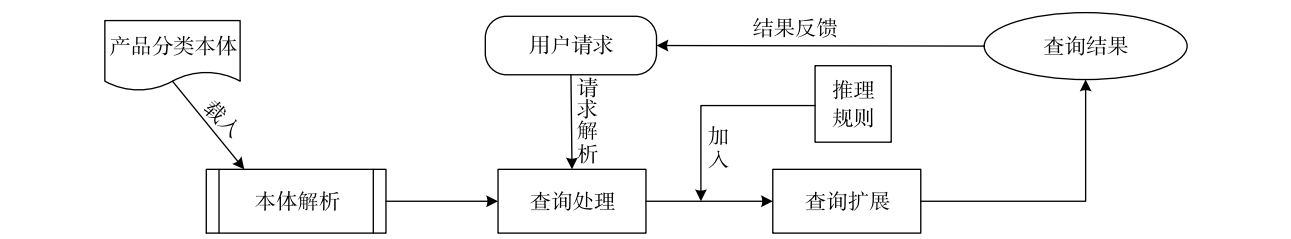


图 5 产品分类的本体查询模型图



图 6 本体解析流程图

表 1 常见本体语言 Jena 解析器

语言	可用解析器	存储方式
OWL-Lite	OWL_LITE_MEM、OWL_LITE_MEM_TRANS_INF、OWL_LITE_MEM_RULE_INF 和 OWL_LITE_MEM_RULES_INF	内存
OWL-DL	OWL_DL_MEM、OWL_DL_MEM_RDFS_INF、OWL_DL_MEM_TRANS_INF 和 OWL_DL_MEM_RULE_INF	内存
OWL-Full	OWL_MEM、OWL_MEM_TRANS_INF、OWL_MEM_RULE_INF、OWL_MEM_MICRO_RULE_INF 和 OWL_MEM_MINI_RULE_INF	内存
DAML + OIL	DAML_MEM、DAML_MEM_TRANS_INF、DAML_MEM_RDFS_INF 和 DAML_MEM_RULES_INF	内存
RDFS	RDFS_MEM、RDFS_MEM_TRANS_INF 和 RDFS_MEM_RDFS_INF	内存

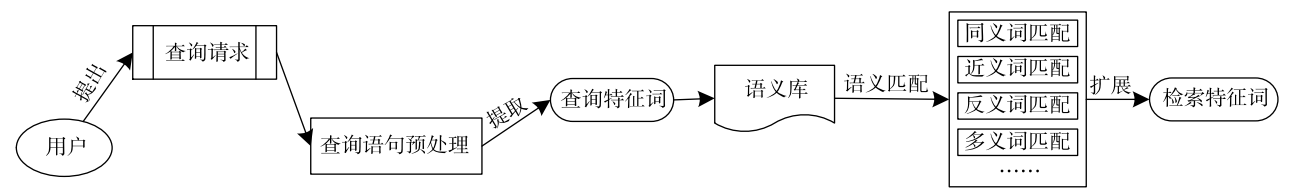


图 7 用户请求模块流程图

- ◆ 本体读取:指通过本体加载模型,将本体文件加载到对应模型,以便进一步进行分析;
- ◆ 本体解析:指的是在本体读取的基础上,通过本体解析器依照本体的五元说,对本体的类、属性、关系、公理和实例进行解析,在属性的处理上区分数据属性和对象属性。常见的本体解析器有 OWL、RDFS 和 DMAL + OIL 等,其中 OWL 又可分为 Lite、DL 和 Full 三个子语言,所用适用子语言和 Jena 解析器如表 1 所示^[25,26]。
- ◆ 解析数据存储:指的是将本体解析的结果,按照一定的数据结构存储。主要包括:对类、对象属性和数据属性的存储、类与类间层级关系存储、属性

与属性间层级关系存储、类与属性间关系存储、公理和实例的存储。

4.1.2 用户请求模块

用户请求模块指的是对用户提出的查询请求进行预处理,并匹配查询特征词。该模块的主要作用是防止用户对产品本体内容结构的不了解,导致的查询命中率的降低。该模块主要包括查询请求、查询语句预处理和特征词语义扩展匹配三个部分,主要流程如图 7 所示。

- ◆ 查询请求:指用户根据自身需求,以自然语言或半结构表达的方式提出的请求。

◆ 查询语句预处理:指将用户提出的查询请求结构化;然后根据构建的产品分类本体形式表达原则,对检索词予以形式化,进行形式匹配。其中形式匹配包括大小写匹配和特殊符号匹配等(如在使用 protégé 构建的本体中,空格分隔符默认为“_”)。

◆ 特征词语义扩展匹配:指对预处理后提取的查询特征词,利用语义库进行语义匹配,得到检索特征词的过程。语义库包括 FrameNet、WordNet 和 VerbNet 等成熟的语义型词典;语义匹配指利用语义库,对查询特征词进行同义词、近义词、反义词和多义词等的匹配,将一个特征词扩展为几个特征词的组合。

4.1.3 查询处理模块

查询处理模块指将用户请求处理模块处理得到的检索特征词与产品本体解析模块解析的数据进行匹配,利用 Jena 推理机和 SparQL 查询器实现用户对查询请求的处理。该模块包括特征词解析、推理机查询和特征集输出三个部分,主要流程如图 8 所示。

◆ 特征词解析:指利用查询处理器对检索特征词进行解析,转换为机器可以理解的表达形式。该阶段的主要对检索特征词进行解析,为用户请求模块提供了分析用户请求的接口。

◆ 推理机查询:指将查询处理器的解析结果载入 Jean 推理机,使用 SparQL 查询器对本体解析模块解析存储数据进行语义匹配的过程。

◆ 特征集输出:指对语义匹配的结果进行获取,得到待输出特征集,准备输出。

4.1.4 查询扩展模块

查询扩展模块指在查询处理模块的基础上,通过往推理机内加入自定义规则,使 Jena 推理机推理能力得到加强,对用户请求处理模块的待输出特征集的查询结果进行语义扩展。包括对查询特征词的间接子类、间接父类、同位兄弟类和该类目下所有的类的 4 个查询扩展。直接子类和直接父类查询在

Jena 推理机中可以直接通过 `getSubClass()` 和 `getSuperClass()` 方法加以实现,不需规则的加入。查询扩展模块弥补了查询处理模块的不足,实现了产品分类从类到其上下所有层级类的查询,同时这些类的层级数亦是可获取的。该模块的查询结果,可以用来进一步分析类目在上下类目中的语义完备性。

4.1.5 推理规则模块

推理指发现概念模型中隐含的内在关系,是类、属性及其之间关系不断丰富过程。一个推理规则一般由大前提、小前提和结论三部分组成,这种结构类似于传统逻辑学中的三段论,在满足大前提和小前提的条件下可推出对应结论,即大前提 \cap 小前提 \rightarrow 结论。大前提指的概念一般范围较广,而小前提的概念一般较窄,两者亦不绝对,在某些情况下可以对调。例如对于推理规则 $[rule: (A \text{ hasSister } B), (A \text{ hasCousin } C) \rightarrow (B \text{ hasCousin } C)]$, A, B, C 是 $Person$ 的三个实例, $hasSister$ 和 $hasCousin$ 是定义的对象属性,对于该规则大前提是 $(A \text{ hasSister } B)$ 或 $(A \text{ hasCousin } C)$,小前提是 $(A \text{ hasCousin } C)$ 或 $(A \text{ hasSister } B)$,结论是 $(B \text{ hasCousin } C)$ 。

推理规则按照其复杂程度可以分内置规则和自定义规则。内置规则为 OWL、RDFS 和 DAML 等语言所特有,其使用比较简单,只需按照固定的语法结构直接调用;自定义规则则是用户根据对应语言的语法规则所自定义的,可以灵活的构造各种复杂关系,并且这种规则亦可单独保存。本模型的推理规则指的是自定义规则。自定义规则按 OWL 已有语法和 OWL 以外的语法来分,可分为描述逻辑自定义规则和其他自定义规则。描述逻辑自定义规则指的是采用集合论和描述逻辑的相关语法结构来描述的规则,如采用 OWL 的 $rdfs:subClassOf$ (子类关系)、 $owl:equivalentClassOf$ (等价关系)、 $rdf:type$ (实例关系)、 $owl:disjointWith$ (互斥关系)等关系进行描述逻辑自定义规则的定义。其他自定义规则采用的语法更为灵活,可以采用特定语境下的对象属性,如

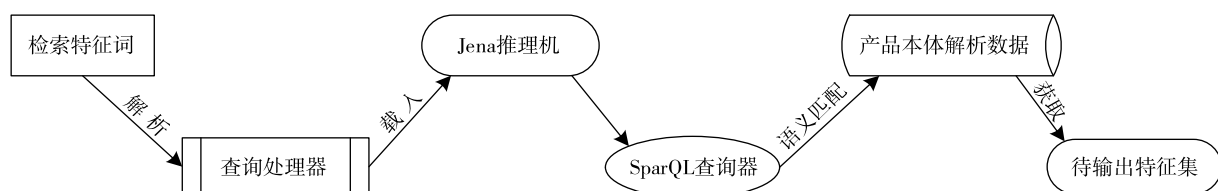


图8 查询处理模块流程图

上文例子定义的 owl: ObjectProperty hasCousin 和 hasSister,这种规则的关系更为松散和贴近实际应用,但带来的后果是覆盖面窄,与特定领域本体的推理挂钩,可复用程度低。因此本模型着重对描述逻辑自定义规则和与产品分类本体查询推理有关的其他自定义规则进行总结,得出如表 2 所示的自定义规则^[27-31]。这些规则可以灵活的组合,应用于产品分类本体查询之中。

表 2 自定义规则

推理规则	形式化表达	规则释义
描述逻辑自定义规则		
rdfs:subClassOf	$(\text{? a rdfs:subClassOf ? b}), (\text{? b rdfs:subClassOf ? c}) \rightarrow (\text{? ardfs:childClassOf ? c})$	a 是 b 的直接子类,b 是 c 的直接子类,则 a 是 c 的子类
	$(\text{? a rdfs:subClassOf ? c}), (\text{? b rdfs:subClassOf ? c}), \text{notEqual}(\text{? a ? b}) \rightarrow (\text{? a info:brotherClassOf ? b})$	a,b 是 c 的直接子类,且 a 与 b 不相等,则 a 与 b 是同位类
rdfs:subPropertyOf	$(\text{? a rdfs:subPropertyOf ? b}), (\text{? b rdfs:subPropertyOf ? c}) \rightarrow (\text{? ardfs:subPropertyOf ? c})$	a 是 b 的直接子属性,b 是 c 的直接子属性,则 a 是 c 子属性
	$(\text{? b rdfs:subPropertyOf ? c}) (\text{? a rdfs:subPropertyOf ? b}) \rightarrow (\text{? ardfs:subPropertyOf ? c})$	
	$(\text{? a rdfs:subPropertyOf ? c}), (\text{? b rdfs:subPropertyOf ? c}), \text{notEqual}(\text{? a ? b}) \rightarrow (\text{? a info:brotherPropertyOf ? b})$	a,b 是 c 直接子属性,且 a,b 不相等,则 a 与 b 是同位属性
owl:DatatypeProperty	$(\text{? a rdfs:subClassOf ? b}), (\text{? x owl:DatatypeProperty ? b}) \rightarrow (\text{? x owl:DatatypeProperty ? a})$	a 是 b 的子类,x 是 b 的数据属性,则 x 也是 a 的数据属性
owl:equivalentClassOf	$(\text{? a rdfs:subClassOf ? b}), (\text{? b rdfs:subClassOf ? a}) \rightarrow (\text{? a owl:equivalentClassOf ? b})$	a 是 b 的子类,b 是 a 的子类,则 a,b 是等价类
owl:equivalentPropertyOf	$(\text{? a rdfs:subPropertyOf ? b}), (\text{? b rdfs:subPropertyOf ? a}) \rightarrow (\text{? a owl:equivalentPropertyOf ? b})$	a 是 b 的子属性,b 是 a 的子属性,则 a,b 是等价属性
rdf:type	$(\text{? a rdfs:subClassOf ? b}) (\text{? x rdf:type ? a}) \rightarrow (\text{? x rdf:type ? b})$	a 是 b 的子类,x 是 a 的实例,则 x 是 b 的实例
owl:disjointWith	$(\text{? a owl:disjointWith ? b}), (\text{? x rdf:type ? a}), (\text{? y rdf:type ? b}) \rightarrow (\text{? x owl:differentFrom ? y})$	a,b 是不相交类,x,y 分别是 a 和 b 的实例,则 x 与 y 不相同
其他自定义规则		
part-of	$(\text{? a info:part-of ? b}), (\text{? b info:part-of ? c}) \rightarrow (\text{? a info:part-of ? c})$	a 是 b 的一部分,b 是 c 的一部分,则 a 是 c 的一部分
kind-of	$(\text{? a info:kind-of ? b}), (\text{? b info:kind-of ? c}) \rightarrow (\text{? a info:kind-of ? c})$	a 是 b 的一种类型,b 是 c 的一种类型,则 a 是 c 的一种类型
part-of 与 kind-of	$(\text{? a info:part-of ? b}), (\text{? b info:kind-of ? c}) \rightarrow (\text{? a info:part-of ? c})$	a 是 b 的一部分,b 是 c 的一种类型,则 a 是 c 的一部分
	$(\text{? a info:kind-of ? b}), (\text{? b info:part-of ? c}) \rightarrow (\text{? a info:part-of ? c})$	a 是 b 的一种类型,b 是 c 的一部分,则 a 是 c 的一部分
accessory-of	$(\text{? a info:accessory-of ? b}), (\text{? b info:accessory-of ? c}) \rightarrow (\text{? a info:part-of ? c})$	a 是 b 的配件,b 是 c 的配件,则 a 是 c 的一部件

4.1.6 查询结果反馈模块

查询结果反馈模块指的是将查询结果可视化地显示给用户,是查询模型的最后一个环节。目前已有的产品分类查询的结果显示较为单一,显示的内容缺乏层次,用户获取查询结果的导出形式有限,不能支持多种方式。该模块包括:查询结果的组织,查询结果的层次化显示和查询结果的多种格式获取。查询结果组织是指将查询结果采用一定的信息构建方法进行组织,适用于显示结果较多的查询;查询结果层次化显示是将查询结果按照类似于产品分类法的层次结构进行显示;查询结果的多种格式获取是指用户可以根据自身需求将查询的结果存储为.txt,.doc 和.xlsx 等多种形式,方便用户进行下一步的数据处理。

4.2 查询实验

在按照产品分类语义分析模型进行本体构建之后,依照本文提出的产品分类本体查询模型进行了查询实验的代码设计,实现了本体扩展查询,包括查询直接子类、间接子类、直接父类、间接父类、同位兄弟类、类目下所有类的六个子查询。其中查询类目下间接子类的关键代码如下:

```
InputStreamReader isr = new InputStreamReader
(System.in);
BufferedReader br = new BufferedReader(isr);
//用户输入查询请求
System.out.println("Please input a classname:");
String keywords = "";
OntModel wgh_ontmodel = ModelFactory.
createOntologyModel ( OntModelSpec. OWL_ MEM,
null); //本体解析
try {
    wgh_ontmodel.read ( new FileInputStream ( " D:\
productclassification.owl" ), null); //本体载入
} catch ( FileNotFoundException e) {
    System.out.println(e.toString());
}
String prefix = "PREFIX rdf: <http://www.w3.
org/1999/02/22-rdf-syntax-ns#>"
+ "PREFIX owl: <http://www.w3.org/2002/
07/owl#>"
+ "PREFIX xsd: <http://www.w3.org/2001/
XMLSchema#>"
```

```
+ "PREFIX rdf: <http://www.w3.org/2000/
01/rdf-schema#>"
+ "PREFIX info: //所查询模型 URI
< http://www.semanticweb.org/administrator/
ontologies/productclassification#>";
String select = "SELECT ? allchildclass ";
String where = " WHERE { ? allchildclass rdf:
subClassOf info:" + keywords + " }";
String wgh_rule = "[ Rule1:( ? a rdf:subClassOf ?
b), ( ? b rdf:subClassOf ? c) -> ( ? a rdf:
subClassOf ? c) ]"; //设计推理规则
Reasoner wgh_reasoner = new GenericRuleReasoner
(Rule.parseRules(wgh_rule)); //将规则加入本体
模型
InfModel inf = ModelFactory.createInfModel(wgh_
reasoner,wgh_ontmodel);
Query query = QueryFactory.create(prefix + select
+ where);
QueryExecution qe = QueryExecutionFactory.create
(query,inf);
ResultSet results = qe.execSelect(); //执行查询
ResultSetFormatter.out ( System.out, results,
query);
qe.close();
```

以构建的 eCl@ss 产品分类本体举例,在用户查询界面输入“Input_device_for_computer”,执行上述代码,查询得到的结果如图 9 所示:图的右侧是所构建的 eCl@ss 本体在 protégé 中打开的示意图,图的左侧是查询代码的运行结果。

通过六个查询子模块的实现,使用户提问不再限于关键词或分类代码,用户可以用更多的方式提出查询请求,使得“使用过于专业”的问题得到极大的缓解。

5 结果讨论

(1) 产品分类本体的推理可判定性

推理的可判定性指的是在有限的时间内得到查询结果,本查询实验均可在有限的时间内获取实验结果。由于本文使用的 OWL 的 DL 子语言,该语言是基于非循环 TBox 框架,确保了查询过程中不会出现语义循环;并且在对产品进行语义分析时,语义分析的结果亦不允许出现环状结构,确保了查询过程中不会出现结构循环。因此避免了在查询推理



图 9 查询间接子类对比图

过程中的逻辑和结构循环,不管查询是否可得到相匹配的结果,查询程序总可在有限的时间内给出答案,保证了推理的可判断性。因此,查询扩展实验验证了基于语义分析的产品分类本体的推理的可判定性。

(2) 产品分类本体语义的正确性

产品分类本体语义的正确性的验证有两方面涵义。一方面指的是在语义分析阶段,类处理和属性处理的正确性,包括类的划分、类的细化处理,属性与类关系的添加、属性的划分等,强调的是产品分类内在内容的正确性;另一方面指的是在语义分析方法的基础上构建的产品分类的正确性,强调的是构建过程的正确性。内容的正确性与体系内容密切相关,构建的正确性由实验结果可得到体现。扩展实验能顺利推理得到上下层级类和同位类,可说明构建的产品分类本体语义的正确性,从查询出的结果可说明内容的正确性。因此,查询扩展实验验证了产品分类本体概念的正确性。

(3) 产品分类本体的语义的完备性

查询扩展实验可以查询到与所查询类有关直接子类、间接子类、直接父类、间接父类、同位兄弟类和所有子类。产品分类本体的语义的完备性指的是

通过这 6 个扩展查询,查询某类目的下子类、父类和同位兄弟类,通过语义分析判断产品分类的分类是否具有语义完备性。

例如在构建的 GPC 本体中查询概念:Computer_Memory(电脑内存)、Call_Id_Displays(来电显示)、Plotter(绘图仪)、Mouse(鼠标)和 Signage(标牌),查询的部分扩展结果如表 3 和表 4 所示。

从上表可以看出,概念 Computer_Memory(计算机内存)的直接父类是 Computer_Components(计算机部件),一般而言,按照具体用途来划分 Computer_Components 可以划分为 Computer_Processors(计算机处理器)、Computer_Components_Variety_Packs(各种计算机组件包)、Computer_Cooling(计算机冷却器)、Computer_Motherboards(计算机主板)、Expansion_Board_Bards(扩展版卡)、Computer_Casing(计算机外壳)、Computer_Components_Other(其他计算机部件)和 Computer_Mass_Storage(计算机大容量存储器)(见蒋本珊《计算机组成原理》2008 版第一章),与实验结果 Computer_Components 同位兄弟类的结果是相一致的,Computer_Memory 按照具体用途的划分结果亦与查询结果相一致。从上表可以看出,类 Call_Id_Displays、Plotter 和 Mouse

也与之类似,在此不加赘述。因此,查询扩展实验证明了利用语义分析方法构建的产品分类本体语义的完备性。

(4) 产品分类本体语义的丰富性

查询扩展可推理得到类全部的子类、父类、同位类,这可从某种程度上反映出产品分类本体具有较为丰富的语义关系和强大的推理能力。本实验仅仅利用了产品分类本体部分的语义关系,若使用更多

的语义关系,可进行更为深入的推理,可根据产品属性逆向查找对应的产品,可进行产品自动归类等,这些能力将是目前分类体系无法比拟的。产品分类本体不仅比现有体系具有更加优越的数据组织结构,而且具有目前产品分类体系缺乏的强大语义关系和推理能力。因此,查询扩展实验验证了产品分类本体语义的丰富性。

表 3 GPC 多组概念查询扩展实验结果 (1)

输入概念	直接子类	直接父类	同位兄弟类
Computer_Memory	Dynamic_Random_Access_Memory Single_Inline_Memory_Module Static_Ram Dual_Inline_Memory_Module RAM Extended_Data_Out Unclassified_Computer_Memory	Computer_Components	Computer_Processors Computer_Components_Variety_Packs Computer_Cooling Computer_Motherboards Expansion_Board_Bards Computer_Casing Computer_Components_Other Computer_Mass_Storage
Caller_Id_Displays	Audio_Announcement Audio_Visual_Combination Visual_Display Unclassified_Display	Communications_Accessorier	Digital_Enhanced_Cordless_Telecommunications_Repeaters Mobilphone_Asscessories Communication_Accessories_Varety_Packs Communication_Accessories_Other Telephone_Wipes
Plotter	Ink_Jet_Plotter Unclassified_Plotter Pencil_Plotter Laser_Plotter	Computer_Output_Device	Printer Copier Monitor Digital_Projector Computer_Output_Device_Other
Mouse	Wired_Mouse Unclassified_Mouse Wireless_Mouse	Computer_Input_Device	Video_Camera Fax Scanner Computer_Input_Device_Other Keyboard Card_Reader
Signage	Signs_Unprinted Signs_Preprinted Signs_Combination Sign_Replacement_Accessory Sign_Holders	Communications	Communication_Variety_Packs Mobile_Communication_Devices_Services Communications_Accessorier Fixed_Communication_Devices

表 4 GPC 多组概念查询扩展实验结果 (2)

输入概念	间接子类	间接父类	全部子类
Computer_Memory	无	Computing Computer	Dynamic_Random_Access_Memory Single_Inline_Memory_Module Static_Ram Dual_Inline_Memory_Module RAM Extended_Data_Out Unclassified_Computer_Memory
Caller_Id_Displays	无	Communications	Audio_Announcement Audio_Visual_Combination Visual_Display Unclassified_Display
Plotter	无	Computing Computer	Ink_Jet_Plotter Unclassified_Plotter Pencil_Plotter Laser_Plotter
Mouse	无	Computing Computer	Wired_Mouse Unclassified_Mouse Wireless_Mouse
Signage	Cleaner Sign_Cover Replacement_Power_Supply_Unit Light_Bulb Control_Knob Replacement_Letters Marker Hanger Unclassified _ Sign _ Replacement _ Accessory	无	Cleaner Sign_Cover Replacement_Power_Supply_Unit Light_Bulb Control_Knob Replacement_Letters Marker Hanger Signs_Unprinted Signs_Preprinted Signs_Combination Sign_Replacement_Accessory Sign_Holders Unclassified_Sign_Replacement_Accessory

参 考 文 献

[1] Kim D, Lee S, Chun J, et al. A Semantic Classification Model for E-Catalogs [C]//Seventh IEEE International Conference on E-Commerce Technology (CEC'05). IEEE Computer Society, 2004 ; 85-92.

[2] Lee H, Shim J. Conceptual and formal ontology model of e-catalogs [M]//E-Commerce and Web Technologies. Berlin Heidelberg:Springer Link , 2005 ; 11-20.

[3] Hepp M, Leukel J, Schmitz V. A quantitative analysis of product categorization standards: content, coverage, and maintenance of eCI @ ss, UNSPSC, eOTD, and the

- RosettaNet Technical Dictionary [J]. Knowledge and Information Systems, 2007, 13(1): 77-114.
- [4] Hepp M. Products and services ontologies: a methodology for deriving OWL ontologies from industrial categorization standards[J]. International Journal on Semantic Web and Information Systems (IJSWIS), 2006, 2(1): 72-99.
- [5] Hepp M, Leukel J, Schmitz V. A Quantitative Analysis of eClass, UNSPSC, eOTD, and RNTD: Content, Coverage, and Maintenance [C]//2013 IEEE 10th International Conference on e-Business Engineering. IEEE Computer Society, 2005: 572-581.
- [6] Barbau R, Krima S, Rachuri S, et al. OntoSTEP: Enriching product model data using ontologies[J]. Computer-Aided Design, 2012, 44(6): 575-590.
- [7] Pagoropoulos A, Andersen J A B, Kjar L L, et al. Building an Ontology of Product/Service-Systems: Using a Maritime Case Study to Elicit Classifications and Characteristics [M]//Collaborative Systems for Smart Networked Environments. Berlin Heidelberg: Springer LINK, 2014: 119-126.
- [8] Natale D A, Arighi C N, Barker W C, et al. The Protein Ontology: a structured representation of protein forms and complexes [J]. Nucleic Acids Research, 2011, 39 (Database issue): D539-D545.
- [9] Wei W, Gulla J A. Sentiment learning on product reviews via sentiment ontology tree[C]//Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2010: 404-413.
- [10] Vegetti M, Leone H, Henning G. PRONTO: An ontology for comprehensive and consistent representation of product information [J]. Engineering Applications of Artificial Intelligence, 2011, 24(8): 1305-1327.
- [11] Panetto H, Dassisti M, Tursi A. ONTO-PDM: Product-driven ONTOlogy for Product Data Management interoperability within manufacturing process environment [J]. Advanced Engineering Informatics, 2012, 26(2): 334-348.
- [12] Anutariya C, Wuwongse V. Ontology Design Approaches for Development of an Excise Duty Recommender System [C]//Information Search, Integration, and Personalization: International Workshop, ISIP 2013, Bangkok, Thailand, September 16 - 18, 2013. Revised Selected Papers. Springer, 2014, 421: 119.
- [13] Buranarach M, Ruangrajitpakorn T, Anutariya C, et al. Ontology Design Approaches for Development of an Excise Duty Recommender System [M]//Information Search, Integration, and Personalization. Berlin: Springer International Publishing, 2014: 119-127.
- [14] Mungall C J, Bada M, Berardini T Z, et al. Cross-product extensions of the Gene Ontology [J]. Journal of biomedical informatics, 2011, 44(1): 80-86.
- [15] Rosati R, Franconi E. Generalized ontology-based production systems [C]//Rome, Italy: Thirteenth International Conference on the Principles of Knowledge Representation and Reasoning. 2012.
- [16] Huang Y, Chen C H, Khoo L P. Product's classification in emotional design using a basic-emotion based semantic differential method [J]. International Journal of Industrial Ergonomics, 2012, 42(6): 569-580.
- [17] Beneventano D, Magnani S. A framework for the classification and the reclassification of electronic catalogs [C]//Washington: ACM Symposium on Applied Computing, 2004: 784-788.
- [18] Borsato M. Bridging the gap between product lifecycle management and sustainability in manufacturing through ontology building [J]. Computers in Industry, 2014, 65 (2): 258-269.
- [19] Sorathia V, van Sinderen M, Pires L F. Towards a Unifying Process Framework for Services Knowledge Management [J]. Lecture Notes in Business Information Processing, 2010(53): 295-299.
- [20] GS1. GPC Publication Browser (01 Dec 2013: Browser-based format, includes translations) [EB/OL]. [2014-11-12]. <http://www.gs1.org/>.
- [21] eCl@ss 6.2. The eCl@ss standard [EB/OL]. [2014-11-12]. <http://www.eclass.de/>.
- [22] Stolz A, Rodriguez-Castro B, Radinger A, et al. PCS2OWL: A Generic Approach for Deriving Web Ontologies from Product Classification Systems [M]//Berlin: The Semantic Web: Trends and Challenges. Springer International Publishing, 2014: 644-658.
- [23] Ramond C K, Assael H. Models of Buyer Behavior, Chapter 19: An Empirical Framework for Product Classification [M]. Clairmont: Marketing Classics Press, 2011.
- [24] Kiang M Y, Ye Q, Hao Y, et al. A service-oriented analysis of online product classification methods [J]. Decision Support Systems, 2011, 52(1): 28-39.
- [25] Horridge M, Bechhofer S. The owl api: A java api for owl ontologies [J]. Semantic Web, 2011, 2(1): 11-21.
- [26] Carroll J J, Dickinson I, Dollin C, et al. Jena: implementing the semantic web recommendations [C]//Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters. ACM, 2004: 74-83.

[27] Grau B C, Halaschek-Wiener C, Kazakov Y. History matters: Incremental ontology reasoning using modules [M]. Berlin Heidelberg:Springer Link, 2007. j. issn. 1000-0135. 2008. 06. 005

[28] Mark Kaminski, Yavor Nenov and Bernado Cuenca Grau. Datalog Rewritability of Disjunctive Datalog Programs and Its Applications to Ontology Reasoning [C]. Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, 2014.

[29] 聂卉. 基于规则的智能检索系统推理机制的研究 [J]. 情报学报, 2008, 27 (6) : 832-838. DOI:10. 3969/

[30] 徐国虎, 许芳, 董慧等. 基于语义关系的本体推理规则研究——以国共合作领域本体库为例 [J]. 中国图书馆学报, 2007, 33 (5) : 88-92. DOI: 10. 3969/j. issn. 1001-8867. 2007. 05. 026.

[31] 董慧, 徐雷, 王菲等. 语义分析系统研究 (I)——史籍语义分析流程 [J]. 情报学报, 2014, 33 (2) : 183-194. DOI:10. 3772/j. issn. 1000-0135. 2014. 02. 008.

(责任编辑 赵 康)