

《智能信息处理》课程考试

基于本体的数字化家具信息检索研究

杨勇超

考核	到课[10]	作业[20]	考试[70]	课程成绩[100]
得分				

2020 年 11 月 30 日

基于本体的数字化家具信息检索研究

杨勇超

(大连海事大学 信息科学技术学院, 大连 116026)

摘 要 随着互联网的快速发展, 网络上遍布了各种信息, 为了有效的获取网络中有价值的信息, 检索技术成为关注的重点。信息检索系统作为网络组成的一部分, 在满足用户信息需求的过程中, 占据着非常重要的位置, 然而, 这种检索方法的查全率和查准率都不能满足用户的需求。为了解决这种问题, 语义网的概念被提出, 因此, 语义检索成为了研究的热点。本文运用语义网中的本体理论, 主要介绍了本体的一些相关知识, 并简单介绍了语义网的概念与体系结构, 最后以数字化家具模型为切入点, 对基于语义网和本体技术的数字化家具信息检索进行研究设计。

关键词 语义网; 本体; 信息检索; 数字化家具模型;

Research on Digital Furniture Information Retrieval Based on Ontology

Yang Yongchao

(School of Information Science and Technology, Dalian Maritime University, Dalian 116026)

Abstract With the rapid development of the Internet, there are all kinds of information on the network. In order to obtain the valuable information effectively, retrieval technology has become the focus of attention. As a part of the network, information retrieval system occupies a very important position in the process of meeting users' information needs. However, the recall and precision of this retrieval method can not meet the needs of users. In order to solve this problem, the concept of semantic web is proposed. Therefore, semantic retrieval has become a research hotspot. In this paper, using the ontology theory of semantic web, this paper mainly introduces some related knowledge of ontology, and briefly introduces the concept and architecture of semantic web. Finally, taking the digital furniture model as the breakthrough point, the research and design of digital furniture information retrieval based on Semantic Web and ontology technology is carried out.

Key words Semantic web; Ontology; Information retrieval; Digital furniture model;

1 引言

随着科学技术的创新和突破, 传统检索方式已无法应对网络海量数据资源, 公众对智能化网络的需求日益迫切, 越来越多的研究者们将目光投向由 W3C 组织制定并推动的语义网, 微软、IBM、HP、斯坦福大学、曼彻斯特大学等大型企业和高等教育机构对语义网技术进行了深入分析研究, 并开发了一系列如 Jena、Racer、KAON 等语义网技术开发应用平台与基于语义网技术的信息集成及查询、推理和本体编辑系统。同时, 我国对于语义网技术的研究也非常重视, 国内已有一批重点高校设立了语义网技术研究中心, 其中, 清华大学研发了 SWARMS 语义网辅助本体挖掘系统, 上海交通大

学研发了 ORIENT 本体工程开发平台等^[1]。目前存在的一些信息检索技术基本上都是基于关键词的, 用这种方法检索出的结果中往往含有许多用户不需要的信息, 准确率比较低; 对于查全率, 这种检索方式不对同义词和相关类等进行处理, 所以也不能保证较高的查全率。这就要求将检索方式从关键字层面提升到语义层面。将信息和语义结合起来, 变成计算机可识别的知识, 从而使系统自动识别出用户的需求, 并挖掘出与用户查询相关的一些信息或者一些用户可能感兴趣潜在信息, 将它们一并返回。

本体是一种可以对概念进行建模的工具, 能够在语义层面描述信息, 同时也可以很好的描述概念的内涵以及概念与概念之间的关系, 具有很好的概念层次, 支持逻辑推理, 因此在信息检索特别是语

义检索领域得到了普遍应用。针对传统检索方式的种种缺陷,本文以家具模型为切入点,将语义检索与数字化家具模型相结合,通过语义网和本体技术对数字化家具模型进行本体构建,研究其关于智能化语义检索的实现方法,提高用户更高针对性的家具信息查询以及提高搜索信息的准确性。

2 研究背景

2.1 语义网

W3C 是语义网提供的一个允许在应用程序、企业和社区边界之间共享和重用数据的公共框架。因此,语义网被视为跨越不同内容、信息应用程序和系统的集成者。语义网与 Web3.0 紧密结合,其作用是连接相关事件及实体。

语义网的 7 层框架体系结构由 Berners-Lee 在 2000 年提出,各层功能自下而上逐渐增强。其中第 1 层到第 4 层分别为用于描述资源位置的字符集层、根标记语言层、资源描述框架层以及本体层,而 5 至 7 层分别为 Logic 逻辑层、Proof 证明层以及 Trust 信任层,Logic 逻辑层为语义本体提供语义公理及推理规则,用于验证资源之间的相互关系以及推理所得结果的有效性;Proof 证明层包含推理过程,为 Logic 逻辑层的规则提供认证;Trust 层提供信任机制,确保信息在网络中进行交换的安全性及可靠性,并且验证该信息是否符合用户要求^[1]。语义网 7 层框架体系结构如图 1 所示。

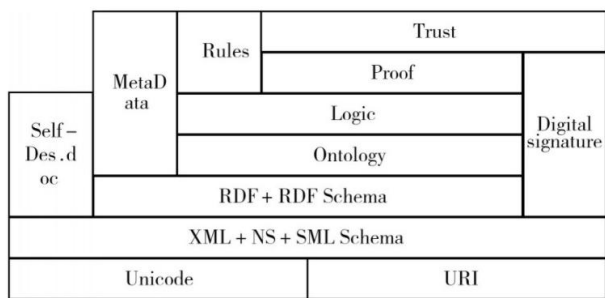


图 1 语义网层次体系

2.2 本体论

本体论起源于西方哲学,是一种探索世界根基的哲学理论。随着信息科学领域的飞速发展,本体的概念被应用于信息科学领域并且被学者们不断完善。从其内涵来看,本体普遍被认为是同一领域内不同主体间相互交流的一种语义基础,指对某一

特定领域的知识进行捕获,并对其进行概念分析、建模,从而对该领域包含的知识进行总结,并将其抽象为一系列共享概念,以得到这些概念在不同层次形式化及模式上的明确定义和描述,最终得到该领域内的共同认可。

本体以一种明确的、形式化的方式表示领域概念及其之间的关系,成为人、机器、应用程序对概念语义达到共同理解的媒介,在应用间实现知识的共享及重用。其获得承认最广泛的定义为 1993 年美国 Sanford 大学的知识系统实验室(KSL, Knowledge systems Laboratory)的学者 ThomasGruber 给出的定义。表述为:“知识的形式化表达的基础是概念化,概念化包括研究领域内的对象、概念和其他实体,以及他们之间的关系;因此,概念化是我们为了某些目的用来表示世界的一种抽象、简化观点,每一个知识库、知识系统、基于知识层次的 Agent 都明确的或是隐含的遵守某个概念化。本体论是对概念化明确的规范说明,在 AI 领域,存在是可以表示。当领域知识以一种明确的形式化进行描述的时候,被表示的对象的集合构成论域,对象的集合及其可描述的关系通过知识表达语言的词汇描述。因此,在 AI 领域,我们可以通过定义只是表示的术语集合来定义程序的本体。在这样的一个本体中,论域中实体的名字通过定义与人可以理解的文本相关联,定义描述了名字包含的意义、公理,公理限制了这些术语的解释及形式化使用。从形式化角度来说,本体是一个逻辑理论的陈述性描述。”

2.3 本体的描述语言

本体描述语言实质上是一种可以对本体进行描述的代码,有助于更好的描述领域内的本体。本体语言要满足以下要求:对语法及语义要有很好的定义、要支持逻辑推理、要具备一定的表达能力、要方便用户使用。目前本体的语言很多,主要可分为 4 大类:非形式化语言、半非形式化语言、半形式化语言、形式化语言。早期的本体语言有 RDF (S)、SHOE、OIL、DAML-ONT、DAML+OIL 等。下面主要介绍 RDF (S)。RDF 即资源描述框架[2],它通常被成为一种“语言”,实际上是一个数据模型(Data-Model)。RDF 对语义的描述比较充分,可借助简单的二元关系描述相关资源信息。RDF 的基本概念是:资源、属性、书目本体构建及语义检索和陈述。资源可以看作是一个对象,任何被描述的事物都是一个资源,如作者、书籍、出版社等。任

意一个资源都能够用一个唯一的 UIR 来标识；属性是一类描述资源之间的关系的特殊资源，也用 UIR 标识，从而避免了分布式数据的命名冲突；陈述是对资源的所有属性进行描述，一个陈述就是描述一个特定资源特定属性的值，是一个形如“对象—属性—值”的三元组（Triple）。每个三元组由三个独立的部分构成：主语、谓语和宾语，分别与对象、属性、值相对应。在 RDF 中任何资源之间的关系都可以用这样的三元组来描述。

3 本体在数字化家具模型中的应用

3.1 数字化家具模型概念及逻辑分析

由于人类自然语言表述的复杂性和不确定性，为使计算机理解自然语言，必须用形式化语言对自然世界进行抽象表示。家具模型根据其作用功能的不同大致可分为床、桌、椅、柜等，根据它们的概念定义对其进行语义描述，如满足人类日常睡眠需求的家具为床，其平面可用于摆放各类物品的家具为桌，供人类坐于其上休息的家具为椅，用于存放物品的家具为柜。不管是哪类家具，基本均由柱类、板材等组件组合而成，这些组件在家具中的作用各有不同，其中柱类家具大多起支撑作用，板材家具中不仅能起到支撑作用，有些还可放置物品。起到支撑作用的柱类、板材一般位于放置物品的板材之下，通过描述这些组件逻辑上的空间位置形成家具整体。在构建家具本体时，根据上述分析对家具模型资源进行语义标注，将语义信息赋予家具资源对象，使其能够被机器理解。

该系统的架构模型在逻辑上分为资源层、语义描述层、本体层、检索层和服务层 5 层，其中资源层的主要作用是搜集各类信息资源并对其进行整合；语义描述层主要利用 RDF 资源描述框架和 OWL 语言对资源进行语义标注；本体层负责本体建立和存储管理，并将构建完成的本体存入本体库内；检索层负责对本体库内存储的本体文件进行解析，并分析用户输入的查询语句，实现语义层面的检索；服务层为用户提供检索结果的可视化界面。该系统架构逻辑模型如图 2 所示。

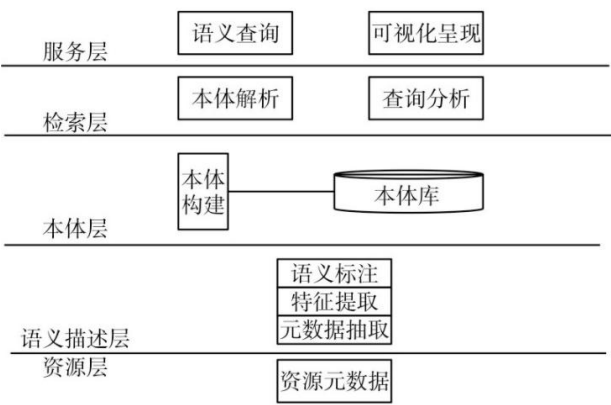


图 2 系统架构逻辑模型

3.2 数字化家具素材模型本体设计

对家具领域的重要术语进行列举，定义类与类的层次体系、属性及属性分面，最后通过本体描述语言创建本体并将本体模型存入数据库从而形成模型素材本体库。通过语义网络规则语言制定相应的本体推理机制。当用户输入查询语句后，由机器对其进行数据预处理操作，即对查询语句按照词语进行划分，并对无用的词语进行清洗。

本文在限定本体构建领域的基础上，确定本体类的构建领域为家具；对本体构建目的和用途进行设定，从而达到增强本体针对性的效果。本文采用斯坦福大学医学院开发的七步法作为本体构建方法，最终得到的语义本体框架如图 3 所示。

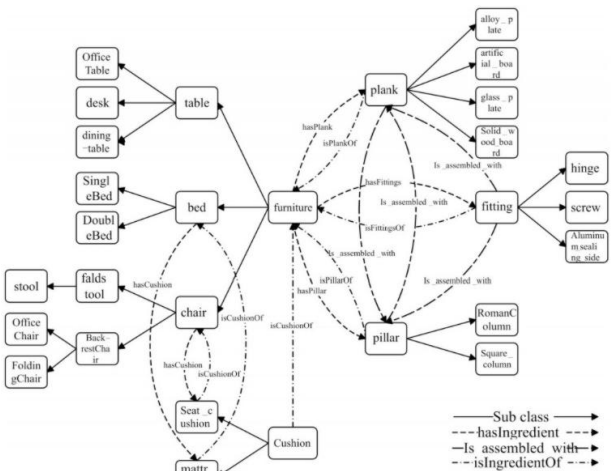


图 3 语义本体框架

3.3 数字化家具本体类定义

面向创意家具素材模型设计的语义本体分别分选取家具（furniture）、垫子（cushion）、柱子（pillar）、板材（plank）作为实体类，每个实体类均可进一步被细化划分为更小的类，形成本体的层次结构。家具实体类可以进一步被细化划分为床

(bed)、桌(table)、椅(chair)、柜(cabinet),而床、桌、椅、柜又可以被进一步划分为更小的概念,如单人床(single bed)、双人床(double bed)、靠背椅(back-restchair)、无靠背椅(faldstool)、办公桌(office table)、餐桌(din- ing table)、衣柜(wardrobe)、书柜(bookcase)等。类的定义如图 4 所示。

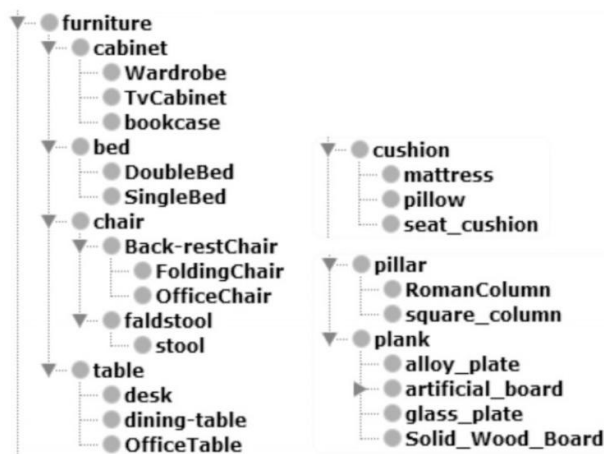


图 4 类的定义

3.4 数字化家具本体类属性及其约束

对类的属性进行定义的过程在整个本体构建过程中至关重要,其可用于描述概念的内在含义,通过复用现有类的属性、扩展或自定义相关元数据标准定义类的属性,同时还需考虑到属性类型,如整数型、浮点型、文本等。在定义类属性的同时,还要对属性的函数性、自反性、传递性等进行约束设定。

针对创意家具素材模型库,对应素材应设定相应数据属性,如坐标(coordinate)、尺寸(size)、材质(texture)、颜色(color)等。针对创意家具模型素材,由于家具是由各种部件拼接组装而成,这些部件是家具类的一部分,但不属于家具类的子类,因此将垫子、柱子、板材这 3 类设定为家具类的兄弟类,与家具类存在包含和部分的关系,如家具类包含垫、柱、板 3 类,垫、柱、板 3 类又是家具类的一部分,同时这 3 个类之间又存在着装配关系。

3.5 基于本体的数字化家具信息检索

通过对本体的创建和管理,逐步扩充和完善本体库,为语义检索提供依据,因此确定研究领域、找出领域内的关键词、依据关键词创建本体、对本

体文件进行存储与维护。针对于数字化家具的小型检索系统,资源比较单一,并不涉及很大的领域或者网络资源,所以构建一个本地的本体就可以满足需求。针对检索步骤进行设计时,首先构造知识库的推理规则,建立本体的语义推理规则,分析创意设计模型的逻辑对象、逻辑关系、概念体系等,以便实现逻辑概念的相互关联,从而完成基于语义的检索。用户输入查询语句后,由后台对数据进行预处理操作,即对查询语句进行分词处理和词语清洗,通过 SPARQL 语句生成最终的检索式执行查询,并将查询结果输出。

推理功能主要是借助 Jena 来实现的,可以对.owl 类型的本体文件进行解析与推理。推理机的推理功能在 Java 程序中通过调用 Jen 实现,首先,系统要读取本地的本体文件、数据资源以及相关的推理规则,在建模之后,系统根据用户输入的检索项进行推理,依据规则对资源进行语义扩展,之后把相关信息一并输出,返回给用户。

语义标注就是建立知识库中实例与本体中概念之间的联系。因为本体从某种程度上来说也是一个小型的知识库,所以向其中添加实例,并对实例的属性赋值,实质上就是进行语义标注。这些资源对象可以使用 URL 引用机制对其进行标识,通过 RDF 链接机制将彼此进行关联,形成语义元数据关联网络,如图 5 所示。

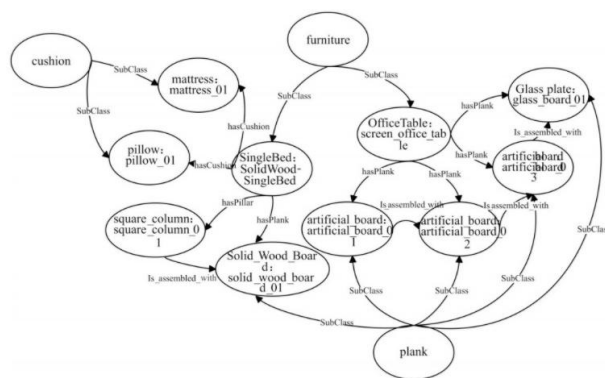


图 5 语义元数据关联

4 结语

本文对基于语义网和本体技术的三维家具素材模型的逻辑概念进行分析,在此基础上设计了语义本体模型,实现了基于语义的检索方法,为构建大众参与的个性化产品协同创意设计平台打下了

基础。如何通过创意设计知识库构建、产品设计要素特征提取与分类、知识迭代与融合及设计交互与智能整合,形成知识与设计的协同演进机制、实现产品个性化定制与创意设计的高效对接是今后研究重点。

参 考 文 献

- [1] 李文靖,胡书山,余日季.基于语义网的数字化家具模型本体设计与检索[J].软件导刊,2019,18(08):136-139+143.
- [2] 刘琪,王小正,王磊.基于本体的教育资源语义检索关键技术研究[J].电脑知识与技术,2014,10(16):3872-3875.
- [3] 王晋,孙涌,王璵玮.基于领域本体的文本相似度算法[J].苏州大学学报(工科版),2011,03:13-17+25.
- [4] 崔韬世,麦范金.词语相似度计算方法分析[J].网络安全技术与应用,2012,05:55-56+72.
- [5] 孙海霞,钱庆,成颖.基于本体的语义相似度计算方法研究综述[J].现代图书情报技术,2010,01:51-56.
- [6] 马良荔,孙煜飞,柳青.语义 Web 中的本体匹配研究[J].计算机应用研究,2017,05:1-3.