

● 颜端武 岑咏华 毛平 成晓

领域知识本体的可视化检索研究*

摘要 军用飞机领域的试验型中文知识本体 OntoAvion 的构建,建立在对国内外典型本体系统、本体构建方法、相关描述语言和构建工具考察研究的基础之上。其检索主界面显示效果可分为4个模块:树形列表检索入口,本体动态图形显示区域,检索控制工具条,检索结果输出区域。它能够将本体中的类层次、属性、实例等语义关系以图形化方式直观显示,实现可视化语义检索。图2。参考文献8。

关键词 领域本体 可视化检索 本体推理 图形交互

分类号 G354

ABSTRACT The authors introduce OntoAvion, a pilot Chinese ontology in the field of military airplanes, which is based on the researches of typical ontology systems and related description languages in China and in foreign countries. Its visualized search interface can be divided into four modules, i. e. a tree-like list, a dynamic graphic ontology display area, a search control tool bar, and a search result output area. It can realize visualized display of hierarchy, attributes, examples and other semantic relations in the ontology. 2 figs. 8 refs.

KEY WORDS Domain ontology. Visualized searches. Ontology reasoning. Graphic interaction.

CLASS NUMBER G354

对领域知识本体的查询检索是开展本体应用的重要环节,其操作一般建立在对本体描述语言和存储文件进行解析和搜索推理的基础之上。典型的本体构建工具如 Protégé, KAON, OntoEdit, OILED 等提供了本体中概念及概念之间联系的描述功能^[1],但在本体显示时多以列表、树状结构或文本超链接方式显示,可视化程度和效果不够,对本体中复杂的网状的语义关系进行动态的可视化表达非常缺乏。本文通过加入可交互的动态网状图形元素,丰富了本体检索的可视化效果。以下将结合面向军用飞机领域的试验型中文知识本体 OntoAvion 的构建工作,重点讨论本体可视化检索的实现效果、采用的实现方案和涉及的关键技术。

1 军用飞机领域知识本体 OntoAvion 的构建简介

军用飞机领域知识本体是对军用飞机学科领域中关于概念体系的明确、正式的规范化说明。该本体主要由军用飞机学科领域知识中的概念、概念之间的关系以及计算机可以识别和处理的形式化描述语言组成。构建军用飞机领域知识本体,可以形成对军用

飞机相关信息组织结构的共同理解,为进一步开展该领域的信息资源整合、语义检索以及跨语种翻译、知识表示、自动标引、知识导航、数据挖掘和知识发现等研究和应用探索奠定基础,具有重要意义。

军用飞机领域试验型知识本体 OntoAvion 构建的工程实践建立在对国内外典型本体系统、本体构建方法以及相关描述语言和构建工具考察研究的基础之上。在构建方法上,我们重点参考了 IDEF5 方法和七步法^[2],并兼顾与国外顶层本体 SUMO 的复用和嫁接问题^[3],在总体策略上采取了自底向上(从底层资源出发)和自顶向下(借助领域专家知识)相结合的思路。主要包括以下几点:①确定本体的专业领域和范畴,即面向军用飞机这一领域的问题,不包括军用直升机、武器火炮等。②考察与国外顶层本体 SUMO 的复用和嫁接问题,采用了与之类似的体系结构,并对嫁接的几个入口进行了分析和判断。③结合领域专家的意见(本课题中我们大量征求了南京航空航天大学两位飞机设计、飞机动力学方面的教授专家的意见),根据军用飞机这一核心概念进行概念延伸,确定其他核心概念,如军用飞机人员、机载武器、军用机场、机载设备、飞机动力系统等。④进行

* 本文受总装备部项目“面向科技数据库网站应用的 Ontology 实体构建研究”(项目号 2005QB1062)和南京理工大学 2005-2006 年度经管青年教师基金资助。

概念填充和关系的迭代,同时添加概念属性的说明,不断进行本体的扩充和完善。关系的确定和概念的选取是本体构建中难度和工作量最大的环节。⑤加入底层概念类的实例,并进行实例属性值的填充。⑥进行一致性检查等。

军用飞机领域试验型知识本体 OntoAvion 的构建工作已历时近一年,工作可分为前期设计和后期存储两个阶段。前期设计中,选用面向对象的统一建模语言(UML),利用 Rational Rose 进行总体设计,后期实际存储工作则采用了 W3C 的推荐标准 OWL 作为本体表示语言,利用 Protégé 工具进行多人协同构建,保证了本体构建的准确性和效率。在概念和关系填充中,参考了《中国图书馆分类法》、《国防科技叙词表》、《航空工业科技词典》、《世界飞机手册》等多部工具书和参考手册,并研读了飞机设计与控制方面的多本教材图书和电子讲义以及 2000 多篇公开发表的军用飞机相关科技论文和 Web 网页,保证了该本体中收录概念的覆盖面、概念和关系的通用性。目前,最新版的 OntoAvion 中已经收录了约 1300 多个概念,包括 400 多个概念类和近 900 个概念实例,以及 70

多个对概念进行说明的属性。这些概念之间的联系通过 12 种抽象型关系来描述,这些关系包括 Protégé 工具中内置的上下位关系 SubClassof、实例同一的关系 SameAs、类-实例的关系 Individual,以及课题组定义的表示定性归属的关系 Belongsto、装备关系 Equips、部分组成关系 isPartOf 和 HasPart、操纵控制关系 Uses、类近义关系 SynSetof、类同一的关系 SameSetof、应用目的关系 HasPurpose 以及用于语言学的词汇组配关系 Facetof。限于篇幅,本文不对该本体的细节进一步介绍。

2 本体可视化检索的实现效果演示

2.1 树形列表检索入口

本体可视化检索主界面(如图 1)左边的树形列表展示本体库的概念层次关系,根结点代表整个军用飞机领域本体。除根结点以外,每个结点都代表本体中的一个概念类,类与子类关系用目录树层级结构显示,文件夹图标表示该类拥有子类,文件图标表示该类除 owl:Nothing 外没有其他子类,图标右边的标签

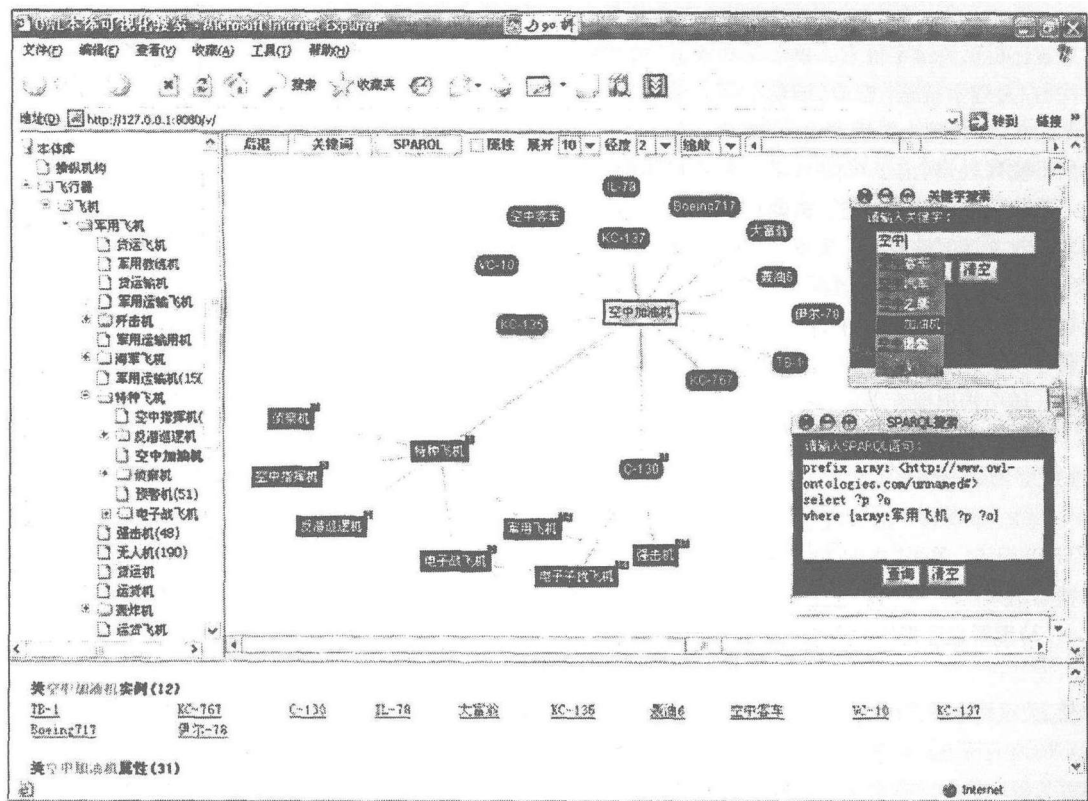


图1 本体可视化检索主界面

值表示该类在本体中的本地名称,括号内的数字表示该类拥有的实例数目。点击树形列表的某个结点,该结点将高亮显示,同时在右边的图形显示区域中将以此结点为中心,动态重绘一个图形界面。

2.2 本体动态图形显示区域

图形显示区域提供本体的可视化导航检索功能。图形由一系列的结点和有向边组成,每个结点代表一个概念类或概念实例,结点的标签值表示类名或实例名,类与实例用不同的形状区分开来,其中直角矩形代表类,圆角矩形代表实例。单击或双击某一结点时,该结点获得焦点并高亮显示,用以区别其他结点,同时将以该结点为中心,绘制出新的导航图。

有向边用来表示类与类之间、类与实例之间、实例与实例之间的语义关系。鼠标滑过某一边时,该边也将高亮显示,并在边的中部绘制出该边的标签值,指示该边代表的语义关系。

2.3 检索控制工具条

检索控制工具条以浮动工具条的方式呈现于图形显示区域的上方,提供关键词检索入口,SPARQL检索入口以及图形显示规模和大小的控制参数设置。

(1)关键词检索入口。除了提供可视化检索,本系统还提供关键字检索。单击工具条上“关键字”将弹出“关键字搜索”窗口,该窗口拥有输入提示和自动完成功能:键入关键词时,在文本框的下方即刻弹出下拉列表,列出了可能与用户输入相匹配的词汇。目前系统只提供关键字精确匹配搜索,可以说提示功能对关键词检索有至关重要作用,因为它除了提示用户输入,另一个重要作用就是能够规范用户输入的词汇。提示的词汇来自本体中的类名、实例名、属性名等。输入关键词后,单击搜索按钮,即可在结果显示区域中输出搜索结果。

(2)为了支持用户自定义搜索,系统还提供SPARQL检索入口。方法是单击工具条上的“SPARQL”,弹出SPARQL检索子窗口,在文本域中输入SPARQL语句,单击“查询”按钮后将在结果区域中输出搜索结果。

(3)图形显示规模和大小的控制参数设置。控制参数设置包括是否显示属性、概念结点展开时的边数阈值、关联概念展开时搜索的边的路径长度以及显示图形的缩放调整滑块等。选中“属性”复选框时,鼠标悬浮位置的概念结点会弹出该概念的相关属性列表。展开阈值默认为10,指定该值后,拥有边数大于该阈值的概念结点初始显示时不展开,直到用户点击。路

径长度决定了关联概念搜索时的深度,目前最高设置为3。通过滑块移动可调整图形显示的大小缩放。

2.4 检索结果输出区域

用户界面的底部是检索结果输出区域。用户选中快捷菜单的“详细信息”,执行关键词检索和SPARQL检索都将在该区域中输出查询结果。在输出的查询结果中,还可单击其中的超链接继续检索。若超链接的文本值是类名/实例名,单击后将在图形显示区域以该类名/实例名为中心重绘整个图形区域,这样文本搜索和可视化搜索可相互配合,达到同步搜索目的。而若用户搜索的是本体个体,系统还能检索出与该个体相关的文献并在结果输出区域显示。

3 本体可视化检索的实现方案与关键技术

3.1 实现方案

本体可视化检索实现方案如图2所示。该系统基于Java开发平台,采用B/S结构,使用Eclipse3.2作为集成开发工具,以Tomcat5.5作为Web服务器和Java Servlet容器,集成了多个Java开源组件。

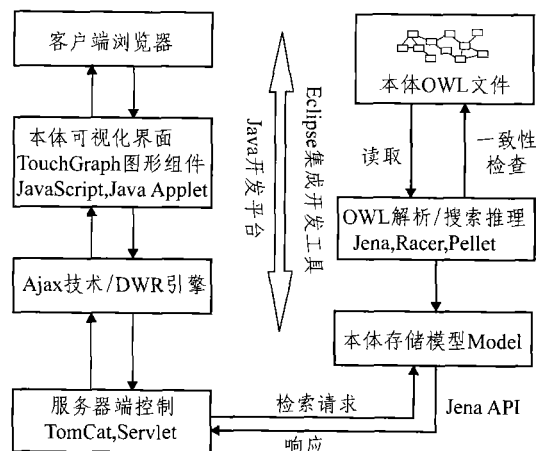


图2 本体可视化检索实现方案

该方案的处理思路为:首先通过Jena, Racer/Pellet等支持OWL解析的API完成对任一个本体资源文件(OWL文档)的内部结构解析;然后根据解析结果将数据(一个个三元组)输出到一个XML文件,输出的文件包含了本体内部结构的相关信息,包括类、属性、实例以及它们之间的语义关系;最后利用TouchGraph应用程序接口,将本体中的类和实例绘制成结点,其中的语义关系绘制成边,从而以可视化的方式把OWL本体表达的语义信息展示给用户,实现本体的可视化检索。

3.2 关键技术说明

3.2.1 服务器端处理

服务器端由一个 Servlet 控制器负责监听所有客户端请求,主要是来自客户端的本体查询请求,Servlet 将请求转交给图 2 中的本体存储模型 Model 类。Model 类是系统的核心类,它通过 Jena 插件^[4]和 Racer, Pellet 推理机^[5]封装了 OWL 文档解析、SPARQL 查询^[6]、本体搜索推理、关键词查询检索等功能。客户端用户在可视化浏览本体的过程中,通过鼠标单击图形或键入关键词将本体查询请求提交服务器,服务器则通过搜索推理引擎和查询引擎得到查询结果,该结果以 XML 的数据形式通过 Servlet 控制器返回给客户端,供客户端显示区域动态刷新时的数据调用。

3.2.2 客户端处理及其与服务器端的交互

客户端浏览器提供了本体可视化检索界面,包括一个树形列表、基于 TouchGraph 组件^[7]的图形交互区域、超链接文本结果显示区和一个浮动的检索控制工具条。军用飞机本体 OntoAvion 的 OWL 文件存在于服务器端,可视化检索界面首次从服务器端向客户端浏览器加载时,会同时生成一个 OWL 解析后输出的 XML 结果文件。该结果文件被一个封装了 TouchGraph 组件的 Java Applet 程序完全加载到客户端内存,并以可视化图形的方式绘制出核心概念“军用飞机”结点以及经度值内的其他关联概念结点。同时,一段 JavaScript 程序也根据该 XML 结果文件绘制出本体结构的树形列表。

客户端程序中还采用了 Ajax 技术和 DWR 引擎^[8],所有查询请求都通过 Ajax DWR 引擎与服务器端交互,实现客户端与服务器异步通信。DWR (Direct Web Remoting) 是一个开源 Java 项目,利用它可以帮助开发人员快速开发使用 Ajax 技术。该 Ajax DWR 引擎在客户端和服务端交流的过程中担负着中间层的任务。用户界面要向服务器提交数据时, Ajax DWR 引擎负责收集数据并通过 Http Request (XML Http Request) 向服务器发送请求,服务器处理完成后返回 XML。Ajax DWR 引擎将 XML 处理为便于用户界面显示的 XHTML 和 CSS 数据,并刷新用户界面相应部分的显示,而非刷新整个页面,避免了不必要的数据往返。结果表明,整个系统就像是一个 Windows 桌面程序一样,没有明显的抖动效果且响应性很高。

3.2.3 可视化检索中的控制与同步

在可视化检索交互中,图形的绘制需要一个合理

的大小和规模,太小则信息量少,太多则会繁杂,并分散用户注意,其控制基于“经度”和“展开边数”两个参数的设置。因为该可视化检索界面提供了多个检索查询入口,所以还要保证各个显示区域的数据同步。

用户关心并选择的当前结点为中心结点,在图形区域中高亮显示。从中心结点出发到其他结点所经过边的条数称为经度(路径长度)。图形显示的层次由经度控制:经度为 0 时,可视化图形只显示中心结点;经度为 1 时,除中心结点外还显示与中心结点直接相连的结点;经度为 2 时,还将显示与中心结点次相邻的结点。可视化图形经度最好控制在 3 以内,超过 3 时,界面上的边将显得过多过杂,太多的边容易分散用户注意力。用户浏览本体时,可根据需要调整经度值。本体中的知识内容呈网状结构,如果不对本体可视化的边进行过滤,整个图形界面就会布满各种各样的边,太多的边会导致一些相对次要的边干扰相对重要的边。因此有必要对图形上的边进行控制,决定是否展开某一结点边。系统对于拥有超过 n 条边的结点默认不予展开,这个 n 值是用户可控制的。同时,用户可自由选择隐藏或显示图形上的任一结点及边。

本体可视化系统提供多个检索入口,包括树形列表检索入口、图形界面检索入口、关键词检索入口以及用户点击操作。树形列表展示本体内部的类层次关系,它实际上向用户提供了对整个本体的全览。图形界面检索由客户端的一个 Applet 实现,用户通过鼠标改变图形中心结点达到可视化浏览本体目的。用户在树形列表选中单击具体的一个概念时,图形显示区域将以该概念为中心重绘整个图形。用户使用关键词检索功能时,若键入的关键词是类名或实例名,客户端向服务器提交查询的同时,图形显示区域也将以该类名或实例名为中心重绘整个图形。这样树形列表查询、文本关键词检索、可视化图形等可相互配合,达到同步搜索目的。

4 结论

本可视化检索系统能够将本体中的类层次、属性、实例等语义关系以图形化方式直观显示,实现可视化语义检索,在此基础上又支持关键字和 SPARQL 检索,同时系统生动的动态二维网状图形也给用户带来了不一样的视觉效果。本研究的主要工作包括:

(1) 针对本体查询和搜索推理问(下转第 76 页)

用户界面建设工作还要与电子政务门户网站建设有机地结合起来。要通过不断完善政府门户网站的设计,增强其智能化性能,方便用户访问。要设置电子政务信息资源主题地图利用指南,解答用户疑问,让用户更加便利地利用政务信息资源。

3.6 维护更新电子政务信息资源主题地图

日常的运行过程中,仍需要对电子政务信息资源主题地图不断进行维护更新。这是因为:第一,主题地图是一个活的东西,会随着应用领域所包含的信息(知识)的进化而进化^[16]。人们对主题的理解会不断更新,以往的某些主题可能不适合要求。这就得进行调整、更改或重新进行设计。第二,在日常运行中,电子政务信息资源系统也会出现一些问题,如数据丢失等。这些都离不开日常维护工作。第三,主题地图虽然能够较好地解决政务信息资源组织中存在的缺乏语义等问题,但在知识组织与表示方面依然存在一些缺陷^[17]。随着理论研究的深入和技术的进步,主题地图的性能必将越来越先进。这势必也要求电子政务信息资源主题地图随技术的进步而不断升级。

维护和更新工作是电子政务信息资源组织系统的一项不可或缺的重要工作,它是主题地图构建工作的延伸。通过维护、更新等措施,不断完善电子政务信息资源主题地图的性能,让它真正成为信息、知识组织和导航的有力工具,优化政务信息、知识结构,方便用户利用。

(上接第63页)题,利用 Jena 插件和 Racer、Pellet 本体推理机,在此基础上设计了 OWL 推理检索引擎,完成了军用飞机领域本体一致性检查和推理检索。

(2)解决了本体可视化递增性浏览难题,动态绘制了本体内部结构和包含的语义关系,实现本体的二维网状的图形可视化检索,动态图形浏览器支持全览、聚焦、移动等功能。

(3)给出了基于 B/S 结构和 Java 平台的本体可视化检索系统的实现方案,可将本体与现有的互联网资源整合,容易使用和共享。

(4)系统提供的关键词检索、树形列表检索、可视化图形检索等多个检索过程同步,使得系统的互动性更高,使用界面更友好方便。

参考文献

- 1 杜小勇,李曼,王大治. 语义 Web 与本体研究综述. 计算机应用,2004(10)
- 2 李景,孟连生. 构建知识本体方法体系的比较研究. 现代

参考文献

- 1,17 韩泽春. 主题地图——一种有效的知识组织与揭示方法. 忻州师范学院学报,2005(4)
- 2 Michel Biezanski, Martin Bryan, Steve Newcomb. ISO/IEC FCD 13250: 1999-Topic Maps. [2006-05-23]. <http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0058.htm>
- 3,7 秦铁辉等. 信息时代的“全球定位系统”——主题地图. 江西图书馆学刊,2005(1)
- 4,8 Steve Pepper. The TAO of Topic Maps. [2006-08-20]. <http://www.ontopia.net/topicmaps/materials/tao.html>
- 5 马费成,郝金星. 概念地图在知识表示和知识评价中的应用(1)——概念地图的基本内涵. 中国图书馆学报,2006(3)
- 6 Kal Ahmed and Graham Moore. An Introduction to Topic Maps. [2006-07-06]. <http://www.architecturejournal.net/2005/issue5/Jour5Intro/>
- 9,10,12,13,15 吴笑凡等. 基于主题地图的异构知识集成. 现代图书情报技术,2005(11)
- 11,16 彭志远. 基于主题地图的知识组织研究与应用[硕士论文]. 兰州大学,2004
- 14 张佩云等. 主题地图标准及其应用研究. 安徽大学学报(自然科学版),2004(3)

吕元智 上海师范大学人文与传播学院讲师,武汉大学信息管理学院2006级博士研究生. 通讯地址:上海市桂林路100号. 邮编200234。

王心裁 武汉大学信息资源研究中心专职研究员,教授. 通讯地址:武昌珞珈山. 邮编430072。

谭必勇 武汉大学信息管理学院2005级博士研究生. 通讯地址同上。(来稿时间:2006-11-02)

图书情报技术,2004(7)

- 3 IEEE Working Group. Suggested Upper Merged Ontology (SUMO). [2006-03-20]. <http://www.Ontologyportal.Org/>
- 4 颜端武,丁晨春等. 基于语义 Web 和 Jena 插件的语义检索系统实验研究. 情报理论与实践,2006(3)
- 5 Pellet OWL Reasoner. [2006-03-20]. <http://www.mindswap.org/2003/pellet>
- 6 SPARQL Query Language for RDF. [2006-04-10]. <http://www.w3.org/TR/rdf-sparql-query>
- 7 TouchGraph Development. [2006-03-10]. <http://touchgraph.sourceforge.net>
- 8 Ajax: A New Approach to Web Applications. [2006-02-08]. <http://adaptivepath.com/publications/essays/archives/000385.php>

颜端武 岑咏华 在读博士,南京理工大学经济管理学院信息管理系讲师. 通讯地址:南京. 邮编210094。

毛平成 晓 南京理工大学经管学院信息管理系情报学硕士生. 通讯地址同上。(来稿时间:2006-11-15)