

基于本体的装备领域信息集成研究

李晓辉,王盼卿,王寅龙

(军械工程学院 计算机工程系,河北 石家庄 050003)

摘要:装备信息集成中存在着大量的语义异构数据源,阻碍了信息的共享和交换。本体可以描述信息之间隐含的关系,在语义和知识层次上描述信息系统的概念模型,有效地解决信息集成中的语义异构问题。通过对基于本体的信息集成方法的分析,采用 Wrapper/Mediator 架构,提出了基于混合本体的装备基础信息集成框架,给出了框架的层次结构和关键技术。结合实际对基于本体的装备指标体系构建和查询处理问题进行了研究,验证了基于本体的信息集成方法在装备领域信息集成中的可行性和有效性。

关键词:本体;信息集成;语义异构;装备领域

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2011)04-0060-05

Research of Information Integration Based on Ontology in Equipment Domain

LI Xiao-hui, WANG Pan-qing, WANG Yin-long

(Department of Computer Engineering, Ordnance Engineering College, Shijiazhuang 050003, China)

Abstract: Large numbers of semantically heterogeneous datasources exist in information integration of the equipment. It baffles the share and exchange of information. Because the ontology can be used to describe the hidden relations between information and describe the concept model of information system on semantic and knowledge. So it can effectively solve the semantic heterogeneity in information integration. On the analysis of ontology-based information integration, puts forward an equipment basic information integration framework based on hybrid ontology, which adopts Wrapper/Mediator structure. Then level and key technique of the framework are given. The construction of equipment Index Architecture and the problem of query processing based on ontology are researched. The feasibility and effectiveness of the information integration method on the basis of ontology are verified in the equipment information integration.

Key words: ontology; information integration; semantic heterogeneity; equipment domain

0 引言

在装备信息化建设中,装备领域信息的管理涉及多个装备业务和管理单位,各装备业务管理部门通常按各自的习惯及工作需要确定数据需求,开发专用的业务信息系统,各信息系统相对比较独立。同时由于缺乏对信息的概念、结构和涵义的统一规范,使得领域内信息在结构、语法和语义方面存在异构,形成了一个相对封闭的“信息孤岛”^[1]。前两种类型的异构可以利用各种中间件技术来解决,但语义异构仍是当前装备领域信息集成中研究的难点。因此需要对装备管理过程中所涉及的信息进行规范,理清信息之间的关系,建立语义数据模型。本体可以客观地描述概念之间的关系,能通过逻辑推理描述概念的语义^[2],因此文

中将本体的理论引入装备领域进行信息集成研究。

1 本体概述

本体(Ontology)最早是哲学上的概念,定义为“对世界上客观存在物的系统描述,即存在论”,是客观现实的抽象本质^[3]。当前,本体论的研究逐渐成熟,并应用到语义 Web、智能信息检索、信息集成、知识工程等领域。

在人工智能界,Neches 等人 1991 年将本体定义为“给出构成相关领域词汇的基本术语和关系,以及利用这些术语和关系规定这些词汇的外延”^[3]。在信息集成领域,本体的定义一直存在争议,1993 年 Gruber 采用形式化定义<D,R>结构,D 表示研究领域内概念的集合,R 则表示概念之间关系的集合,并且把 Ontology 解释成“Ontology 是概念模型的明确的规范说明”^[4]。1998 年 Studer 等人认为“本体是共享概念模型的明确形式化规范说明”^[5],该定义被业界普遍接

收稿日期:2010-08-25;修回日期:2010-11-29

基金项目:全军装备信息化建设“509 工程”

作者简介:李晓辉(1978-),男,讲师,研究方向为信息与系统集成;
王盼卿,教授,研究方向为 C³I、信息与系统集成。

受,包含4层含义:概念化、明确化、形式化和共享。从Ontology的定义可以看出本体的目标是捕获相关领域的知识,提供对知识的共同理解,确定该领域内共同认可的词汇,并给出这些词汇之间的相互关系,因此将本体应用于异构信息集成,具有描述数据源语义和解决语义异构的潜力,具体表现在公共语义描述、基于语义的查询和信息重用三个方面。

2 基于本体的装备基础信息集成框架

本体在信息集中,根据应用领域、目的的不同,主要有3种不同的集成方法:单本体方法、多本体方法和混合本体方法^[6-9]。单本体方法使用一个全局本体为信息集成提供统一的共享词汇库来表示信息的语义,数据源很多时候是跨领域的,要从所有数据源综合抽象出一个全局本体难度很大。在多本体方式中,每一个数据源都建立相应的局部本体描述信息的语义,局部本体之间需要建立映射关系,当数据源较多时不同本体间的映射关系很难定义。混合本体方法结合了以上两种方法的优点,各数据源建立相应的局部本体,为了各局部本体之间可以比较,在局部本体之上建立了一个全局共享词库即全局本体。

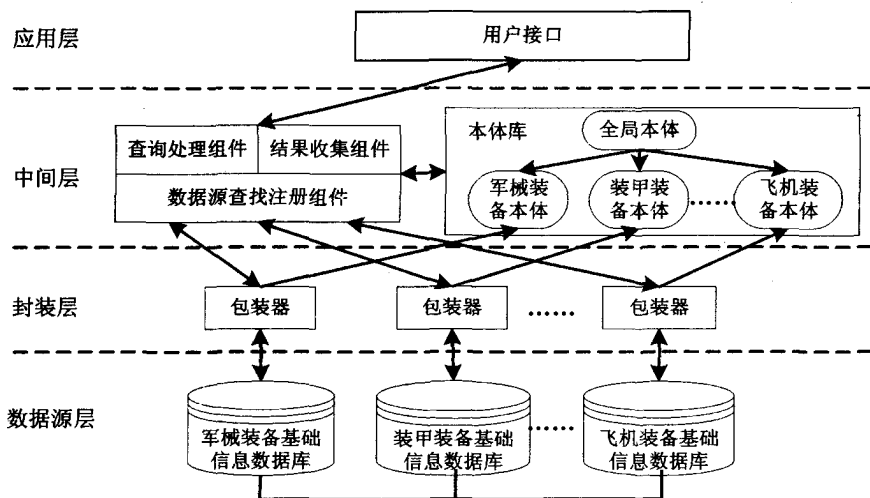


图1 基于本体的装备基础信息集成框架

通过对装备基础信息语义异构性的分析,基于混合本体方法,采用Wrapper/Mediator架构,信息集成框架如图1所示,整个框架分为四层:数据源层、封装层、中间层和应用层^[10]。数据源层处于框架的最底层,是数据的提供者,涉及各个装备大类的基础信息,不同数据源信息的组织方式、数据模型和内涵不尽相同,需要利用本体对信息的概念、关系进行统一,屏蔽异构性;封装层由被称作包装器(Wrapper)的软件模块组成,每个数据源有一个相应的包装器,负责与底层数据源进行通信,数据源在注册的时候,抽取数据源的信息,交给中间层处理,另一方面接收中间层的查询请求从

数据源中抽取数据;中间层用于屏蔽各种数据源的异构性,包括数据源注册组件、查询处理组件、结果收集组件和本体库,查询处理、结果收集组件通过与本体库进行交互完成查询的分解和结果的汇总排序操作;应用层提供统一的用户界面,接收用户的查询请求并提交给中间层处理,另一方面从中间层获取最终查询结果以统一的视图进行显示。

该框架采用了混合本体方法,即对每一个装备基础数据源建立相应的局部本体,然后在此基础上通过比较、分析、合并创建全局本体并建立全局本体与局部本体的映射关系。全局本体为用户提供一个共享词汇库,用户针对全局本体的查询通过全局-局部本体的映射关系转换分解为各个数据源的子查询,子查询结果通过组装合并提交给用户。文中利用该信息集成框架对基于本体装备指标体系构建和查询处理问题进行研究。

3 基于本体的装备指标体系构建

装备基础信息管理中主要的概念有装备、指标和指标数据值。装备之间存在父子、兄弟、整体与部分的关系;指标与装备之间的关系是属性关系,即装备具有

一些特定的指标;指标数据值与指标之间是实例关系,即某一装备具体的指标数据值是该指标的一个实例。在构建装备指标体系时,可以引入本体的概念,描述装备、战技指标,以及装备与战技指标之间的内在关系,进行装备指标体系的合理构建。

3.1 本体结构分析与设计

根据装备目录树的层次结构和所属指标的特点,可以将指标分为通用指标、共

用指标和专用指标,通用指标作为树根的属性,共用指标作为同层兄弟结点中指定结点的属性,专用指标作为树叶结点的专用属性。某一具体装备的指标体系可以根据装备之间的关系进行构建,主要由三部分组成:一是直接从父结点继承,二是从兄弟结点引用,三是本身所具有的专用属性。下面通过具体实例进行分析,图2是一个具体的指标构建本体示例图。

在该本体示例图中,阴影结点为同层兄弟结点中具有共用指标的结点,如“中型坦克”结点;通用指标“生产单位”和“列装时间”为树根结点“装甲装备”的属性,共用指标“最大行程”和“最大射程”为“中型坦

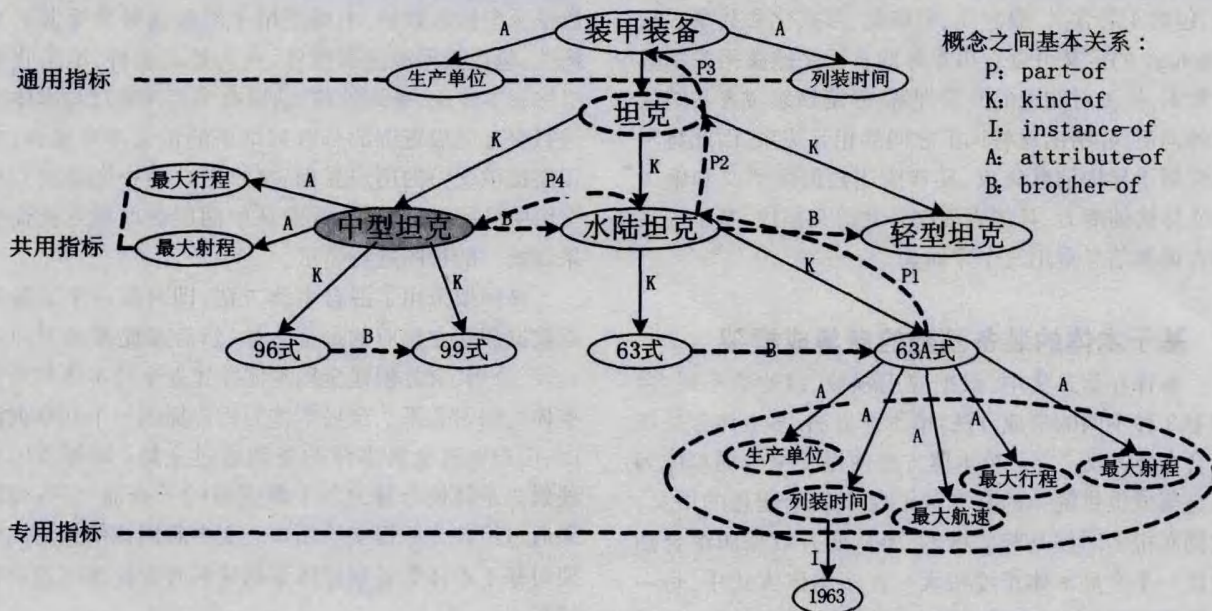


图 2 本体示例

克”结点的属性,专用指标“最大航速”为“63A式”结点的属性;子结点可以从父结点继承属性,兄弟结点可以引用共用指标;根据装备结点之间的父子、兄弟关系,可以合理地构建某一装备结点的指标体系,例如叶子结点“63A式”的战技指标可以通过继承和引用构建,其中“生产单位”和“列装时间”沿着路径{P1,P2,P3}从结点“装甲装备”继承而来;“最大行程”和“最大射程”沿着路径{P1,P4}通过继承、引用关系来自于“中型坦克”结点;“最大航速”属性是其本身所具有的专用指标。

通过以上分析,该本体可以定义为一个四元组:
Ontology = < Concepts, Relations, Rules, Instances >。其中:

Concepts:表示装备分类、装备以及装备战技指标属性名称的集合;

Relations:概念之间关系的集合,包括父子关系、兄弟关系、等价关系、实例关系;

Rules:表示概念和关系之间规则的集合,包括子结点从父结点继承战技指标属性、兄弟结点之间引用战技指标属性两条规则;

Instances:表示装备战技指标属性值的集合,每个装备的每一个属性的值都是该属性的实例。

可以对本体的结构给出形式化的定义,定义如下:

Type

OntologyNode = Record

NodeName:string; //结点名称

NodeId:string; //结点标识,具有层次结构,可以识别父结点和兄弟结点

Bzw:boolean; //是否树叶结点标识

Attr:string; //通过属性的标识识别其所具有的属性集

State:boolean; //结点状态标识,是否为指定结点

Grp: integer; //同层结点的排序

End;

3.2 装备指标体系的构建

通过对本体结构的分析,在构建本体结构树的基础上,可以根据装备之间的关系构建某一具体装备的指标体系。构建时通用指标从父结点继承,共用指标从兄弟结点引用,最后根据指标的分类属性对通用、共用和专用指标进行分类排序构成装备的完整指标体系。采用该方法进行指标体系的构建体现了装备分类之间的各种关系,便于指标的整体维护,同时消除了指标数据项的语义异构性,使得系统具有更强的健壮性,下面给出指标体系构建的形式化算法。

算法:构建指定装备的指标体系

输入:装备本体结点 OntologyNode

输出:指标集合 Attr

{Attr=NULL //初始化指标集合为空

GetAnceNode(OntologyNode) //得到存储指标的祖先结点
集合{O₁,O₂,...,O_n}

For i=1 to n do

AddAttr(O_i,Attr) //继承祖先结点的通用指标,添加到指标集合

GetBrotherNode(OntologyNode) //得到存储指标的兄弟结点
集合{O₁,O₂,...,O_n}

For i=1 to n do

AddAttr(O_i,Attr) //引用兄弟结点的共用指标,添加到指标集合

AddAttr(OntologyNode,Attr) //添加本结点专用指标到指标集合

Class(Attr) //根据指标属性对指标进行分类

Sort(Attr) //根据指标属性对指标进行排序

Return Attr //返回指标集合

4 基于本体的装备基础信息查询处理

查询处理需要解决不同装备大类之间信息在语义上的异构性,通过全局本体与各大类装备本体之间的映射关系进行查询的分解转换,屏蔽信息的异构性,将查询结果以统一的格式进行显示。查询处理包括:提出查询请求、查询分解、子查询执行和生成查询结果四个过程。

4.1 查询请求

用户针对全局本体中的概念和属性提出查询请求,全局本体实际上是一个共享的词汇库,提供对全局概念的统一规范化定义,用户利用全局查询检索各装备大类中的异构信息。全局查询 Q 用类 Sql 语句进行形式化,定义为三元组 $\langle S, F, W \rangle$ [11,12],其中 S 表示 Select,后接用户要查询的变量,为全局本体中概念; F 即 From,表示在哪些局部本体中查询; W 代表 Where,表示查询变量与本体库中其他变量的关系,包括变量之间的关联关系和常量之间的比较关系 ($>$ 、 $<$ 、 $=$ 、 \geq 、 \leq 、 \neq)。

4.2 查询分解

通过全局-局部本体的映射关系,将使用全局本体概念的全局查询重写为针对各个局部本体的子查询。分解后的查询语句用局部本体中的概念和关系进行描述。其中 Select 和 From 子句根据全局-局部本体的映射关系转换,Where 中的常量直接映射,变量如果是同一装备大类数据源则进行映射,否则不做处理。全局查询经过分解,转换为相应装备大类的子查询。

4.3 子查询执行和生成结果

不同装备大类基础信息数据源可能采用不同的查询语言(如 XQuery、SQL 等),通过包装器 Wrapper 进行子查询转换,使得各子查询与对应数据源的查询语言相匹配,然后执行子查询。各子查询返回结果后,由结果收集组件根据全局-局部本体映射表对查询结果进行语义分析、汇总、排序和去冗余操作,进行结果的组装,提交给用户。

4.4 实例分析

下面以具体的实例分析查询处理过程,如图3所示,设“军械”与“装甲”两个局部本体分别为 O_1 、 O_2 ,通过局部本体的合并即形成全局本体 O_g 以及全局-局部本体的映射关系。通过全局本体-局部本体的映射,可消除“军械”与“装甲”装备大类基础信息之间的异构性,映射表存储全局属性与各类装备属性之间的映射关系和映射规则,全局查询通过查找映射表分解成相应装备大类的子查询。查询处理过程分析如下,其中 epname、atname、data 分别表示装备名称、指标属性名称和指标属性数据值。

假定用户对于 O_g 的全局查询为类 Sql 语句:

$Q = \text{Select epname, atname, data From 军械装备, 装甲装备 Where (atname = '参考价格' and data > 12 \text{ 万元}) and (atname = 'MTTR' and data > 2 \text{ 小时})}$ 。

根据 From 子句的装备大类名称以及全局与局部本体之间的映射关系,可以将 Q 分解为子查询 q_1 、 q_2 ,其中:

$q_1 = \text{Select epname, atname, data From 军械装备 Where (atname = '参考价格' and data > 12 \text{ 万}$

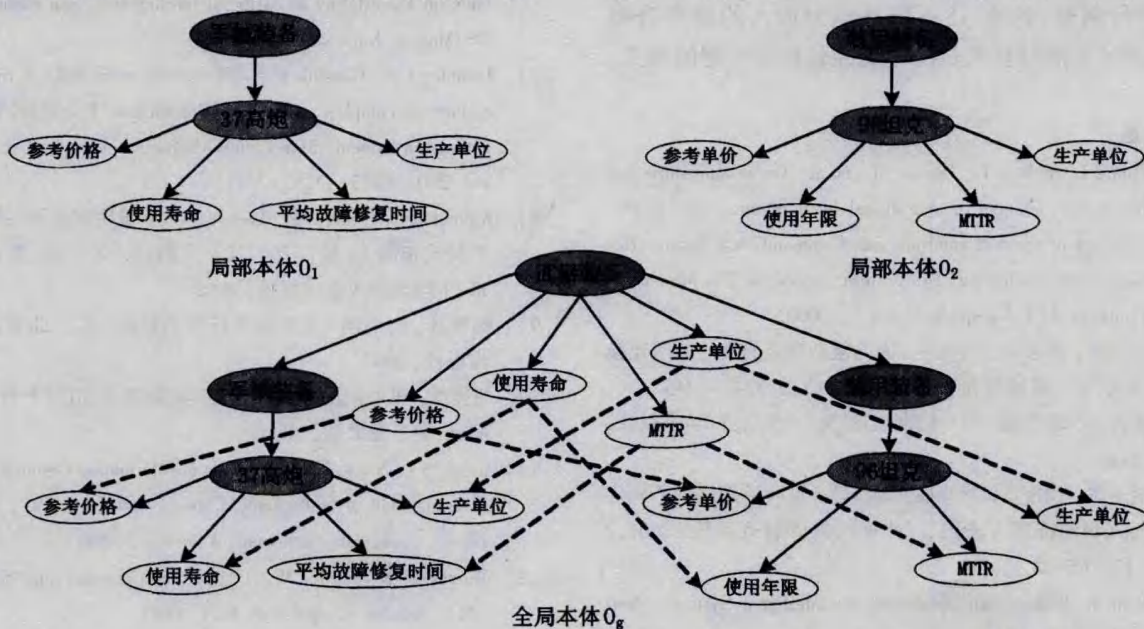


图3 全局-局部本体

元) and (atname = '平均故障修复时间' and data > 2 小时);

$q_2 = \text{Select } \text{epname}, \text{atname}, \text{data} \text{ From 装甲装备}$
Where (atname = '参考单价' and data > 12 万元)
and (atname = 'MTTR' and data > 2 小时);

通过查询分解,解决了“平均故障修复时间”和“MTTR”、“参考单价”和“参考价格”之间的语义异构问题。将子查询 q_1 和 q_2 分别应用于信息源“军械装备”和“装甲装备”大类执行查询,对查询结果进行汇总、排序操作即可完成查询结果的组装集成。

5 结束语

基于本体的信息集成是当前信息集成研究的热点,文中通过对装备领域信息特点的分析,提出了基于全局-局部本体映射的装备基础信息集成框架,并对本体结构进行了分析设计,研究了基于本体的装备战技指标体系构建算法,结合实例对该集成框架中的查询处理问题进行了研究。随着本体在装备领域信息集成的深入研究,利用本体将有效地解决该领域信息集成中的语义异构问题,实现信息的共享和交换。

参考文献:

- [1] 于琦,周勇.一种基于本体的异构数据源模式集成[J].计算机技术与发展,2008,18(2):34-37.

(上接第 59 页)

入,通过加权和方法,得到所需的全部概率分布,保持了概率分布的一致性。同时,分析人员可以根据自己的期望值,通过对相对权值的微小的调整来对概率分布进行调整;此外,还可调整作为输入的概率分布值,这对于发展分析人员的主观经验具有一定的意义。

参考文献:

- [1] Zhang L, Falzon L, Davies M, et al. On relationships between key concepts of operational level planning[C]//Proceedings of the 5th International Command and Control Research and Technology Symposium. Australia War Memorial, Canberra ACT, Australia:[s. n.], 2000.
- [2] 胡笑旋,杨善林,马溪骏.面向复杂问题的贝叶斯网建模方法[J].系统仿真学报,2006,18(11):3242-3246.
- [3] 张连文,郭海鹏.贝叶斯网引论[M].北京:科学出版社,2006.
- [4] 胡玉胜,涂序彦,崔晓瑜,等.基于贝叶斯网络的不确定性知识的推理方法[J].计算机集成制造系统,2001,7(12):65-68.
- [5] Pearl J. Probabilistic Reasoning in Intelligent Systems: Network of Plausible Inference[M]. San Francisco: Morgan

- [2] 吴昊,刑桂芬.基于本体的信息集成技术研究[J].计算机应用,2005,25(2):456-458.
- [3] 邓志鸿,唐世渭,张铭. Ontology 研究综述[J]. 北京大学学报,2002,38(5):730-738.
- [4] Gruber T R. A translation approach to portable ontology specifications[R]. Stanford: Knowledge System Laboratory, Stanford University, 1993.
- [5] Studer R, Benjamins V R. Knowledge Engineering, Principles and Methods[J]. Data and Knowledge Engineering, 1998, 25(1-2):161-197.
- [6] Wache H, Vogele T, Visser U, et al. Ontology-based integration of information—a survey of existing approaches[C]//In Proceedings of IJCA-01 Workshop: Ontologies and Information Sharing. Seattle, WA:[s. n.], 2001:108-117.
- [7] 李星毅,高文浩,施化吉.基于本体的异构数据集成方法[J].计算机工程与设计,2009,30(8):1931-1933.
- [8] 喻洁,夏安邦.基于本体的电力多数据源信息集成研究[J].计算机应用与软件,2009,26(6):124-125.
- [9] 冯志勇,李文杰.本体论工程及其应用[M].北京:清华大学出版社,2007.
- [10] 王晓芳.基于本体的异构数据源集成系统模型及其查询处理[D].济南:山东大学,2006:16-17.
- [11] 李晓辉.基于本体的装备领域信息集成研究[D].石家庄:军械工程学院,2008.
- [12] 李晓辉,王盼卿,齐剑峰,等.基于本体的异构信息查询技术研究[J].军械工程学院学报,2007,19(6):64-66.

Kaufmann Publishers, 1988.

- [6] Diez F J. Parameter adjustment in Bayes networks: the generalized noisy OR-gate[C]//Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence. San Francisco, US: Morgan Kaufmann, 1993.
- [7] Lemmer J F, Gossink D E. Recursive noisy OR—A rule for estimating complex probabilistic interaction[J]. IEEE Transactions on System, Man, and Cybernetics—Part B: Cybernetics, 2004, 43(5):2252-2261.
- [8] Kahneman D, Solvic P, Tversky A. 不确定状况下的判断: 启发式和偏差[M]. 方文, 吴新利, 张擎, 等译. 北京: 中国人民大学出版社, 2008.
- [9] 杨善林, 胡小建. 复杂决策任务的建模[M]. 北京: 科学出版社, 2007.
- [10] 胡笑旋. 贝叶斯网建模技术及其在决策中的应用[D]. 合肥: 合肥工业大学, 2006.
- [11] Amari S I, Nagaoka H. Methods of Information Geometry[M]//Translation of Mathematical Monographs, vol. 191, Rhode Island: American Mathematical Society, 2000.
- [12] Murray M K, Rice J W. Differential Geometry and Statistics[M]. London: Chapman & Hall, 1993.