数字图书馆

编者按：数字图书馆栏目由同方知网技术有限公司协办。同方知网与电子杂志社以网络出版和知识情报服务为主要业务方向，依靠自主开发的全文数据库管理、知识挖掘与数字出版等先进技术，与社会各界通力合作，坚持打造可为全社会提供各种知识服务的《中国知识资源总库》。其中，国家重点出版项目——《中国学术文献网络出版总库》，大规模集成整合了我国学术期刊、博硕士学位论文、会议论文、报纸、年鉴、工具书、学术图书、专利、标准、科技成果等各类文献资源。尤其是基于《总库》的行业、专业与个性化数字图书馆，融合了各类先进的知识服务模式，为高效率创新、学习和管理决策创造了理想的信息化环境。

中文叙词表本体一致性检测机制研究与实现*

曾新红 林伟明
（深圳大学图书馆 深圳 518060）
明 仲
（深圳大学信息工程学院 深圳 518060）

【摘要】研究中文叙词表本体（OntoThesaurus，即基于中文叙词表建立的本体知识库）的一致性检测机制，并将其应用在中文叙词表本体共建共享系统（OTCSS）的修订意见提交、叙词表本体更新和全局检查等相关过程的实现中，取得了良好的应用效果。

【关键词】叙词表 本体 中文叙词表本体 本体构造 一致性检测 本体演化
【分类号】G254 TP311

Research and Implementation of Consistency Checking Mechanism for OntoThesaurus

Zeng Xinhong Lin Weiming
(Shenzhen University Library, Shenzhen 518060, China)
Ming Zhong
(College of Information Engineering, Shenzhen University, Shenzhen 518060, China)

【Abstract】This paper studies the consistency checking mechanism for OntoThesaurus. It is used in OTCSS (OntoThesaurus Co_construction and Sharing System) to realize the submission of amendment opinion, updating and global checking of OntoThesaurus, and works effectively.

【Keywords】Thesaurus Ontology OntoThesaurus Ontology construction Consistency checking Ontology evolution

收稿日期：2008-02-14
收修改稿日期：2008-03-03
*本文系国家社会科学基金项目“基于本体和知识集成实现中文叙词表的升级、共享和动态完善”（项目编号：05CTQ001）和国家自然科学基金项目“视角理论及其在本体集成中的应用”（项目编号：60673122）的研究成果之一。

1 引 言

叙词表是本体出现之前最高端的知识组织系统，其编制有严格的国际和国家标准规范。但由于我国现有的主题词表(叙词表)大部分是手工编纂的，难免出现错误。在将其进行了本体化升级之后，可以借助本体语言的推理能力对其进行严格的一致性检测，理清和补全相关信息，从而建立起严格的概念体系和词间关系体系，极大地提高叙词表的科学性。

中文叙词表本体(OntoThesaurus)是基于中文叙词表建立的本体知识库。中文叙词表本体共建共享系统OTCSS将已有的中文叙词表转换为OWL文件，并通过在其网络共享应用过程中采集使用者知识来实现其动态更新完善。初始OntoThesaurus首先要进行一次全局一致性检测，以修正叙词表中原有的错误。随后，在OntoThe-saurus的共建过程中，一致性检测机制的应用可以大大降低修订意见提交者和修订专家的工作量和错误率，并保证更新后的OntoThesaurus不出现新的一致性冲突。因此，一次性检测机制的研究和实现对OntoThesaurus在整个生命周期中的健康运行至关重要。

2 OntoThesaurus的TBOX构建

OntoThesaurus的类定义和属性定义请参见参考文献[1]中的表1和表2。本文作了进一步的研究，进行了以下修改和扩展：

(1)直接以叙词作为概念的表述形式，取消Term类、PTerm类和HasPTerm属性。此举大大缩小了本体的容量并简化了实现过程。

(2)参考ANSI/NISO Z39.19－2005[2]的第8节(Relationships)，对Broader，Narrower和Related三个属性分别进行了子属性扩展，详见参考文献[3]。此举利于将初始的粗粒度OntoThesaurus逐渐演化为细粒度本体，从而支持基于概念间具体子关系的推理。

3 OntoThesaurus中存在的一致性问题

笔者认为，用于人工智能目的和推理目的的本体应该是严格控制的知识组织系统，其规范程度不应低于叙词表。因此笔者在OntoThesaurus中保留了叙词表结构中的精华(即须符合汉语叙词表编制规则(GB13190－91)[4]中的核心要求)，使叙词表界几十年来在术语控制上的研究成果得以延续。同时OntoThesaurus也是一个本体，必须符合本体理论和技术的要求，这就要求必须将叙词表中供人理解的规则表述明确定义为机器可理解的形式化规范说明。本体技术的应用也允许在保持其功能的前提下抛弃叙词表规则中纯粹为方便人工使用而制定的一些规则(如非叙词款目可不再存在)。

OntoThesaurus的一致性问题具体表述如下：

(1)叙词必须一词一义[4]。

叙词因其在标引和检索中的特定功能而要求绝对单义[4]，这是术语控制的重要一步。

隐含：一个概念在OntoThesaurus中只能由一个叙词来表示；叙词在OntoThesaurus中必须明确定义；入口词不能与叙词同形。

(2)等同关系，指叙词与非叙词之间的关系[4]。

笔者在OntoThesaurus中取消了“等同关系必须是双方互相指引”的规则，即只保留叙词款目，取消非叙词款目，此举对降低系统的实现复杂度有重要意义。检索时同样可从入口词检索到叙词，非叙词的指引作用仍存在。输出书本式叙词表格式时也可由程序自动生成非叙词款目。

(3)属分关系，指上位叙词与下位叙词之间的关系，必须相互指引[4]。

隐含：属分关系是叙词之间的关系；属分关系是互逆的；属分关系是与直接上下位词的关系，不可越级，否则无法生成层次化的词族等级。

叙词款目中显示直接属分关系，有利于使用者通过词间关系明确词义，并可启发读者进行扩检和缩检，笔者认为是合理的冗余，可为系统减少不必要的推理负担。因此在OntoThesaurus中仍规定属分关系必须成对出现。

扩展的属分关系子关系也必须遵守以上规则。

(4)相关关系，指叙词之间属分以外的相关关系[4]。

隐含：相关关系是叙词之间的关系；相关关系不能与属分关系(及其子关系)重合。以上规则同样适用于扩展的相关关系子关系。

(5)为了有效控制OntoThesaurus的规模，避免出现低级错误和不必要的冗余，笔者还明确了以下隐含规则：

所有词间关系都是反自反的,即不能是术语与其自身之间的关系;除相关关系外,其他词间关系都是反对称的,即关系两边的概念不可互换。

上述规则在叙词表编制规则中虽然没有明确提出,但作为一种常识性的共识,在各种叙词表的实际编制过程中已作为一种默认的潜规则而得到严格执行。

4 OntoThesaurus 一致性问题的形式化描述

本节用形式化方法来明确定义上述一致性问题。

4.1 叙词定义缺失问题

在叙词表中,除了等同关系是叙词与非叙词之间的关系外,属分关系和相关关系都是叙词之间的关系,且叙词必须明确定义。相应地,在 OntoThesaurus 中,除了 HasNTerm 以外的其他 ObjectProperty 都是概念(Concept)之间的关系,而 Concept 的实例在 OntoThesaurus 中必须明确定义。其形式化定义为:设有关系 R,若存在 x,y 满足 R(x,y),且 R 的值域为{x|Concept(x)},而在 OntoThesaurus 中未明确定义 Concept(y),则判定 y 缺失定义。

运用"未定义叙词"检测可以查出在叙词款目中作为属、分、参、族等关系词出现,而又未明确定义为叙词的术语。

4.2 值域不一致问题

值域是指属性的取值范围,其形式化定义为:设 C 为本体中的概念,C 有属性 R,则 R 的值域表示为:range(R) = {y|∃x (R(x,y)∧C(x))}。

值域不一致是指知识框架中的属性取值不在定义的值域范围内。其形式化定义为:设 C 为本体 M 中的概念,C 有属性 R,如果 M 中存在 R(c,c'),其中 C(x) 且 c' ∉ range(R),则 M 存在值域不一致。

OntoThesaurus 中包含叙词 Concept、非叙词 NTerm 以及词间的代 HasNTerm、属 Broader、分 Narrower、族 TopConcept、参 Related 等关系。其中 Broader、Narrower、Related(及其子关系)和 TopConcept 等关系(属性)的值域均是{x|Concept(x)}(即它们都是叙词之间的关系),而代关系 HasNTerm 的值域是{x|Nterm(x)},且 Concept∩NTerm =∅。因此 OntoThesaurus 可能存在值域不一致的问题。

运用"值域不一致"检测可以查出入口词与叙词同形的错误(即一个术语既是叙词又是入口词)。

4.3 OntoThesaurus 中的 HasNTerm 关系是反函数型的

反函数型定义如下:设 R 为定义在集合 X 上的二元关系,∀x∀y∀z ((R(x,z)∧R(y,z))→x = y),则称 R 是反函数型的。

在 OntoThesaurus 中规定,代关系属性 HasNTerm 是反函数型的,即一个入口词不能出现在多个叙词之下。形式化定义为:若关系 R 是反函数型的,存在 x,y,z 满足 Concept(x)∧Concept(y)∧Concept(z)∧R(x,z)∧R(y,z)∧x≠y,则 OntoThesaurus 出现一致性问题。

本条规则通过检测入口词的多次出现,可检测出同一个概念在 OntoThesaurus 中出现多个叙词的情况。若确系不同概念使用了同一个入口词,则允许例外(或为入口词添加限定词进行区分)。

4.4 OntoThesaurus 中的所有词间关系是反自反的

为了控制 OntoThesaurus 的规模,保持较小的冗余度,规定其中的所有词间关系(HasNTerm、Broader、Narrower、TopConcept、Related 以及所有扩展的子关系)均应是反自反的(即不能是术语与其自身之间的关系)。

反自反的定义是:设 R 为定义在集合 X 上的二元关系,如果∀x ∈X (x,x) ∉R,则称 R 是反自反的。如果 OntoThesaurus 中的词间关系 R,存在 x 满足 Concept(x)且 R(x,x),那么判定 OntoThesaurus 是不一致的。

4.5 除相关关系(Related)外,OntoThesaurus 中的其他词间关系都是反对称的

反对称的定义是:设 R 为定义在集合 X 上的二元关系,如果∀x∀y ((R(x,y)∧R(y,x))→(x = y)),则称 R 是反对称的。对于 OntoThesaurus 中 Related 之外的关系 R,如果存在 x,y 满足 Concept(x),Concept(y)且有 x≠y,R(x,y),R(y,x),则判定 OntoThesaurus 是不一致的。

运用"非法对称关系"检测可检查出某些低级错误,例如 A 属 B 而 B 又属 A 的情况。

4.6 OntoThesaurus 中任意两个词间关系(TopConcept 除外)之间不能存在同一个二元组

除了 TopConcept 与 Broader 关系可能共享同一个二元组(即一个叙词款目中的属关系词和族首词是同一个叙词)外,OntoThesaurus 的所有词间关系的任意两个关系之间不能存在同一个二元组。形式化定义为:设 R_1,R_2 是 OntoThesaurus 中的词间关系,且 $R_1 \neq R_2$,

如果存在 x,y 满足 Concept(x) ∧ Concept(y) ∧ R_1(x,y) ∧ R_2(x,y),那么 OntoThesaurus 是不一致的。

运用"二元关系冲突"检测可检查出某些词间关系错误,例如 A 分 B 同时 A 又参 B 的情况。

4.7 OntoThesaurus 中互逆的一对关系的断言必须成对出现

逆关系的定义是:设 R 为 X 到 Y 的二元关系,R 的逆关系 $R^{-1}=\{(y,x)|R(x,y)\}$。

在 OntoThesaurus 中,互逆的一对关系(如 Broader 和 Narrower)的断言必须成对出现。也就是说,设 R_1,R_2 为 OntoThesaurus 中的两个关系,且 $(R_1)^{-1}=R_2$,存在 x,y 满足 Concept(x) ∧ Concept(y) ∧ R_1(x,y),如果 $(y,x)\notin R_2$,那么判定 OntoThesaurus 出现信息缺失,须补充缺失的信息。

4.8 OntoThesaurus 中具有传递性的关系的断言不能出现越级

OntoThesaurus 中的属分关系及其子关系是具有传递性的。传递性的定义是:设 R 为定义在集合 X 上的二元关系,如果 $\forall x\forall y\forall z((R(x,y)\wedge R(y,z))\rightarrow R(x,z))$,则称 R 是传递的。而参照叙词表编制标准的规定,在 OntoThesaurus 中这些具有传递性的关系,它们的断言只反映当前叙词上下一级的属分关系,不能出现越级情况,这样才能保证能够根据属分关系推理出严格的词族等级结构。其形式化定义是:在 OntoThesaurus 中,设 R 具有传递性,定义关系 R' 如下:

$$\forall x\forall y(R(x,y)\rightarrow R'(x,y))$$

$$\forall x\forall y\forall z((R'(x,y)\wedge R'(y,z))\rightarrow R'(x,z))$$

如果存在 x,y,z 满足 Concept(x) ∧ Concept(y) ∧ Concept(z) ∧ R(x,y) ∧ R(x,z) ∧ R'(y,z),则判定 OntoThesaurus 存在越级情况,是不一致的。

系统可以根据通过了一致性检测的属分关系推理出严格的词族等级结构,并自动补充族首词。

5 OntoThesaurus 一致性检测机制的实现

笔者运用 Jena 提供的基于自定义规则的推理机实现了 OntoThesaurus 的一致性检测机制。Jena 是 HP 公司的开源语义网应用框架,它为 RDF、RDFS 和 OWL 提供了可编程环境。Jena 的推理机制有 3 种:使用 Jena 自带的基于一般规则的推理机、使用自定义规则的推理机和使用外部推理机。Jena 开发包对 OWL 的推理提供了完备的支持[5]。

5.1 自定义规则

Jena 提供基于自定义规则的推理机[5],它通过一定的推理引擎来解释这些规则,并完成推理。用户可根据需要定制自己的规则,然后创建特定的推理机来完成推理。Jena 的推理机提供前向链、后向链和混合式的推理引擎。其中,前向链和后向链推理引擎可以独立使用,也可以使用前向链引导后向链推理引擎。本文采用前向链推理引擎来实现 OntoThesaurus 的一致性检测机制。前向链推理引擎基于标准的 RETE 模式匹配算法,该算法由 Charles Forgy 博士在 1979 年提出,是在模式匹配中利用推理机的时间冗余性和规则结构的相似性,通过保存中间运算来提高推理效率的一种模式匹配算法。算法的核心思想是对分离的匹配项根据内容来动态构造匹配树,以达到降低运算量的效果[6,7]。

5.2 运用自定义规则检查一致性

在上两节中讨论的 OntoThesaurus 的一致性问题可以通过使用 Jena 的自定义规则推理机来解决。具体的自定义规则如下:

(1)叙词定义缺失问题。解决思路为:

先运用规则[r2:(?x rdfs:subClassOf ?y)(?a rdf:type ?x) → (?a rdf:type ?y)]来补全子类的实例,然后运用规则[r1:(?p rdfs:range pre:Concept)(?x ?p ?y)(?x rdf:type pre:Concept) noValue(?y rdf:type pre:Concept) → (?y rdf:type pre:Concept)(?y pre:err 'error1')]来查出缺失定义的叙词。

(2)值域不一致问题。解决思路为:

先运用规则[r2:(?x rdfs:subClassOf ?y)(?a rdf:type ?x) → (?a rdf:type ?y)]来补全子类的实例,然后运用规则[r12:(?p rdfs:range pre:NTerm)(?x ?p ?y)(?y rdf:type pre:Concept) → (?y pre:err 'error2')]来查出值域不一致的问题。

(3)OntoThesaurus 中的 HasNTerm 关系是反函数型的。解决思路为:

运用规则[r10:(?x pre:HasNTerm ?y)(?ox pre:HasNTerm ?y) notEqual(?x,?ox) makeTemp(?z) → (?z pre:err 'error10')]来查出不一致的地方。

(4)OntoThesaurus 中的所有词间关系均是反自反的。解决思路为:

先运用规则[r2:(?x rdfs:subClassOf ?y)(?a rdf:type ?x) → (?a rdf:type ?y)]来补全子类的实例,然后运用规则[r5:(?p rdf:type owl:ObjectProperty) notEqual(?p,pre:TopConcept) (?x ?p ?y) (?x rdf:type pre:Concept) equal(?x,?y) makeTemp(?z) → (?z pre:err 'error5')]来查出不一致的地方。注意,这里不考虑族首词这个角色,在一致性检查后每个实例的族首词可通过程序自动生成。

(5)除 Related 外,OntoThesaurus 中的其他词间关系都是反对称的。解决思路为:

先运用规则[r2:(?x rdfs:subClassOf ?y)(?a rdf:type ?x) → (?a rdf:type ?y)]来补全子类的实例,然后运用规则[r4:(?p rdf:type owl:ObjectProperty) notEqual(?p,pre:Related) (?x ?p ?y) (?y ?p ?x) notEqual(?x,?y) (?x rdf:type pre:Concept) makeTemp(?z) → (?z pre:err 'error4')]来查出不一致的地方。

(6)OntoThesaurus 中任意两个词间关系(TopConcept 除外)之间不能存在同一个二元组。解决思路为:

运用规则[r6:(?x ?p ?y) (?p rdf:type owl:ObjectProperty) notEqual(?p,pre:TopConcept) (?x ?q ?y) notEqual(?q,pre:TopConcept) notEqual(?p,?q) (?q rdf:type owl:ObjectProperty) (?x rdf:type pre:Concept) makeTemp(?z) → (?z pre:err 'error6')] 来查出不一致的地方。注意,这里不考虑族首词这个角色,原因同(4)。

(7)OntoThesaurus 中互逆的一对关系的断言必须成对出现。解决思路为:

运用规则[r8:(?p owl:inverseOf ?q) (?x ?p ?y) noValue(?y ?q ?x) (?x rdf:type pre:Concept) (?y rdf:type pre:Concept) makeTemp(?z) → (?z pre:err 'error8')]来查出信息缺失的地方。

(8)OntoThesaurus 中具有传递性的关系的断言不能出现越级。解决思路为:

由于属分关系是具有传递性的,且它们互为逆关系,因此在补全了逆关系的情况下只需检查其中一种情况即可。建立一临时的角色 TBroader(代表任意级的上位关系),先运用以下 3 条规则[r9a:(?a ?p ?b) (?p rdfs:subPropertyOf pre:Broader) (?a rdf:type pre:Concept) (?b rdf:type pre:Concept) → (?a pre:Broader ?b) (?a pre:TBroader ?b)],[r9b:(?a pre:TBroader ?b) (?b pre:TBroader ?c) (?a rdf:type pre:Concept) (?b rdf:type pre:Concept) (?c rdf:type pre:Concept) → (?a pre:TBroader ?c)]和[r9c:(?a pre:Broader ?b) (?a rdf:type pre:Concept) (?b rdf:type pre:Concept) → (?a pre:TBroader ?b)]找出所有传递级别的三元组,然后运用规则[r9d:(?a pre:Broader ?b) (?a pre:Broader ?c) notEqual(?b,?c) (?b pre:TBroader ?c) (?a rdf:type pre:Concept) (?b rdf:type pre:Concept) (?c rdf:type pre:Concept) makeTemp(?z) → (?z pre:err 'error9')]来查出越级情况。

(9)此外,还可以利用自定义规则推理为 OntoThesaurus 自动补全族首词。解决思路为:

运用以下规则[r13a:(?a ?p ?b) (?p rdfs:subPropertyOf pre:Broader) → (?a pre:Broader ?b)][r13b:(?a pre:Broader ?b) (?b pre:Broader ?c) → (?a pre:Broader ?c)][r13c:(?x rdfs:subClassOf ?y) (?a rdf:type ?x) → (?a rdf:type ?y)]来补全叙词之间的上位关系(包括间接的上位关系),然后再运用规则[r13c1:(?a pre:Broader ?b) noValue(?b pre:Broader ?c) → (?a pre:TopConcept ?b)]来自动生成各个叙词的族首词。

5.3 一致性检测机制的具体实现

笔者运用 Jena 开发包来具体实现 OntoThesaurus 的一致性检查机制。步骤如下:

(1)运用 Jena 提供的 ModelFactory 来创建一个 Ontology Model,调用该 Model 的 read 方法将 OWL 格式的 OntoThesaurus 读入 Ontology Model 中。

(2)根据具体要求写出对应的推理规则。

(3)运用 Jena 提供的 GenericRuleReasonerFactory 读入推理规则形成推理器 Reasoner。

(4)根据 Reasoner 以及(3)创建的 Ontology Model 来创建推理模型 InfModel。

(5)可调用 InfModel 的 prepare 方法来触发推理规则的运行,也可通过读取推理模型的信息来触发。

6 一致性检测机制的应用效果

目前,OntoThesaurus 的一致性检测机制已应用在中文叙词表本体共建共享系统(OTCSS)的修订意见提交、叙词表本体更新和全局检查等相关过程的实现中,

取得了良好的效果。下面以《敦煌学检索词表本体共建共享系统》为例,着重介绍分步全局检查的步骤和应用效果。在修订意见提交、叙词表本体更新过程中,系统有针对性地运用了以下这些步骤的局部或全部,以减少修订意见提交者和修订专家的工作量和降低他们的出错率,并保证 OntoThesaurus 在整个生命周期中的健康运行。

对于刚刚从已有的中文叙词表转换而来的初始 OntoThesaurus 来说,由于大部分现有中文叙词表是手工编制的,可能存在比较多的一致性问题,可以先按以下顺序逐步检测和排除一致性问题,最后再进行一次批量的全局检查。按以下顺序检测可以使系统和修订专家的工作量最小化,因为前一个问题可能是导致后一个问题出现的重要因素。而对于直接使用 OTCSS 系统建立(即将已有叙词表输入或完全新建)的 OntoThesaurus 来说,由于在建设过程中已经过了比较严格的一致性检测,发生错误的可能性较小,可以在发布共享之前直接进行一次批量的全局检查,以从全局的角度发现和排除最后的错误。

6.1 未定义叙词

"未定义叙词"检测可查出在已有叙词款目的词间关系中出现而又未定义为叙词(即不存在叙词款目)的术语。修订专家可通过网络界面查看相关款目详细信息,判断是否应为其添加叙词款目,然后通过自动添加或修改/删除已有款目相关信息等手段进行处理。

未定义叙词的检测使 OTCSS 系统可以在叙词款目的保存过程中自动添加相关叙词的叙词款目(带有自动生成的拼音和由互逆关系推出的词间关系等信息),从而可大大降低输入工作量,加快建库过程。

6.2 值域不一致的问题

通过值域不一致检测可查出既是叙词又是入口词的术语(见图 1)。例如,如图 2 所示:"维摩经疏"本身是一个叙词,同时它又是另一个叙词"维摩经义疏"的入口词。修订专家可通过网络界面进行修改处理,消除冲突。

6.3 入口词多次出现

通过"入口词多次出现"检测可以检查出以下错误:为同一概念选用了多个叙词,或由于输入错误未能查重而造成同一概念出现多个叙词。此时应只保留一个叙词,其他作为入口词出现或予以删除,如图 3 图 4 所示。

还有一种特殊情况就是,不同的概念(叙词)确实

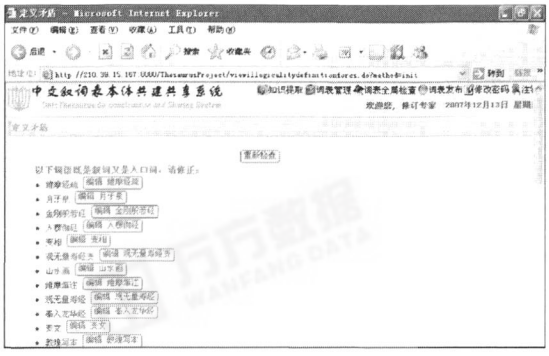

图 1 值域不一致检测结果

图 2 值域不一致示例

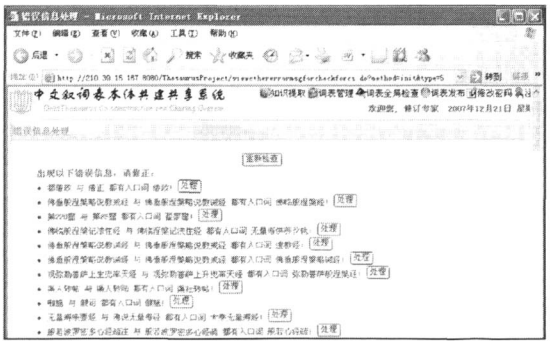

图 3 入口词多次出现检测结果

存在同形的入口词。例如在《敦煌学检索词表》中,古"空谷山"、"三危山"和"无穷山"今均俗称为"火焰山",敦煌石窟代表窟"第 220 窟"和"第 85 窟"亦均称为"翟家窟",经查阅《敦煌学大辞典》[8]无误,建议保留。

6.4 非法自反关系

从逻辑角度来看,自反关系(即术语与其自身之间

图 4　入口词多次出现示例

的关系)是关系的一个特例,在某种意义上不能算是一种错误。但自反现象的存在会徒然增加系统的实现和维护成本,因此在结构严谨的知识组织系统中有必要杜绝这种现象的存在。在传统的中文叙词表的编制过程中其实已默认执行了这条规则,所以这种现象只是偶有发生(《敦煌学检索词表》中仅检出 3 条)。在 OntoThesaurus 的动态完善过程中,非法自反关系的检测可有效控制 OntoThesaurus 的规模膨胀。

6.5　非法对称关系

通过非法对称关系检测可检查出手工编制词表中可能出现的某些低级错误,如两个术语都是叙词而又互为入口词的情况(见图 5 和图 6),或一个叙词与另一个叙词互为上位词等错误。

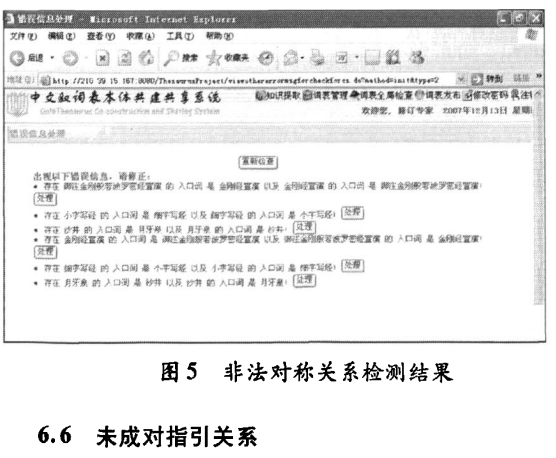

图 5　非法对称关系检测结果

6.6　未成对指引关系

未成对指引关系的检测可检查出必须成对出现的互逆关系的缺失,系统可做自动补齐处理或提示人工处理(见图 7 和图 8)。属分关系未能相互指引是手工编制叙词表中的常见错误,通过这一步检测和修订处理可全部排除。通过 OTCSS 系统对 OntoThesaurus 进

图 6　非法对称关系示例

行动态完善的过程中,因有严格的未成对指引关系检测,可以自动补充添加成对指引关系,从而可杜绝此类错误的再发生。

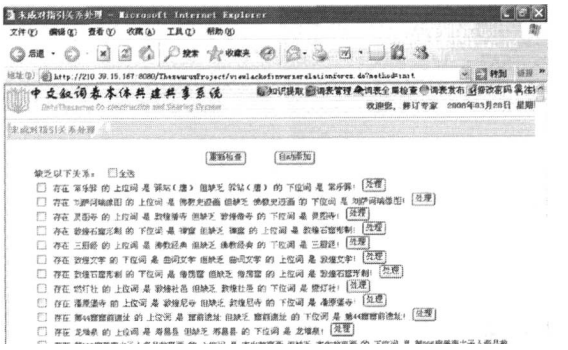

图 7　未成对指引关系检测结果

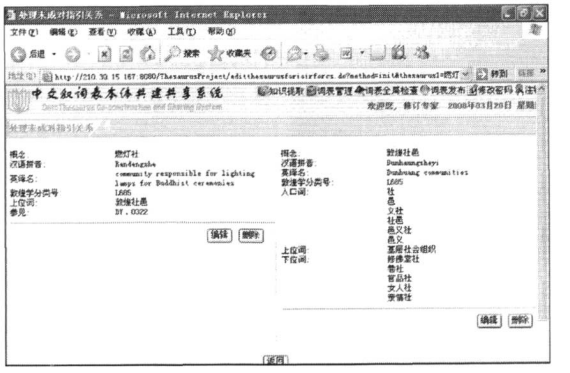

图 8　未成对指引关系示例

6.7　二元关系冲突

在手工编制的叙词表中,二元关系冲突偶有发生。例如,如图 9 和图 10 所示,在《敦煌学检索词表》中出

现的一个二元关系冲突错误:“敦煌俗文学”参“说唱故事”同时又分“说唱故事”。此时修订专家只能选择保留其中一个关系。

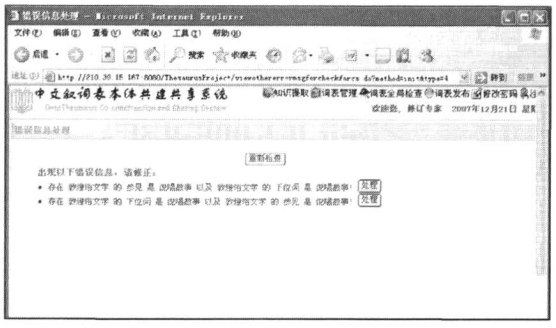

图9 二元关系冲突检测结果

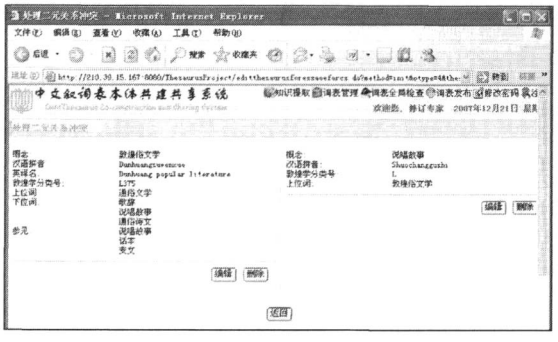

图10 二元关系冲突示例

6.8 传递关系越级

为了获得层次分明的词族等级结构,在 OntoThesaurus 的叙词款目中只能出现最近一级的属分关系(或其子关系)。也就是说,传递关系不可越级。这是手工编制词表较难把握的问题,比较容易出现错误。通过传递关系越级检测可以找出并排除这种错误。例如,如图11和图12所示,在《敦煌学检索词表》中检测到的一个传递关系越级错误:“大辟图”同时属“经变画”和“敦煌壁画”,而“经变画”又属“敦煌壁画”。此时修订专家可通过网络界面比较分析此错误所涉及的3条款目的详细信息,在“大辟图”款目中删除越级的上位词“敦煌壁画”,即可理顺传递关系。

6.9 自动生成族关系

在传统的叙词表修订过程中,需要在后期根据修订过的属分关系人工调整词族等级和族首词。而在

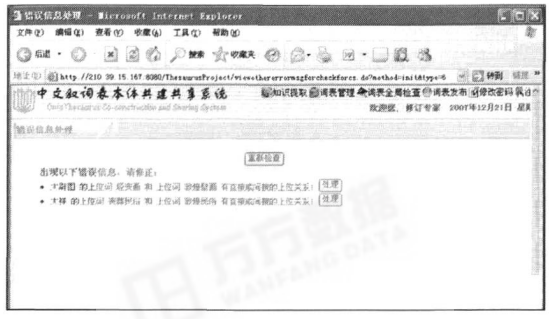

图11 传递关系越级检测结果

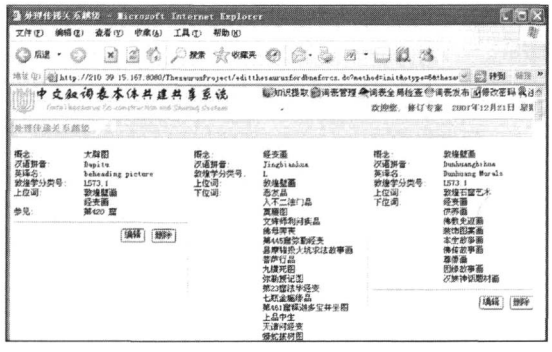

图12 传递关系越级示例

OntoThesaurus 中,这一步可以由机器自动完成。鉴于使用者调用查看整个词族并非高频率事件这个事实,笔者选择采用使用者需要时再通过推理动态生成词族的方式提供服务(《中国分类主题词表》二版电子版也采用这种方式)。而族首词的存在有利于这一服务的实现,因此笔者在全局检查的最后一步为具有属分关系的叙词自动生成族关系(族首词)。

若要求严格按属分关系的各种子关系来生成词族等级,则需要考虑为族关系增加相应的子关系(如类属族首词、实例族首词和整体/部分族首词)。

7 结 语

OntoThesaurus 一致性检测机制的实现有力地保证了中文叙词表本体在其生命周期每个阶段的质量。该机制对其他知识组织系统(如分类法、规范档等)和知识组织系统表示方法(如 SKOS[9])具有较好的可移植性,对实现这些知识组织系统和表示方法的一致性检测具有直接的参考价值。

(致谢:敦煌研究院信息资料中心为本文提供了重要的研究实

例——《敦煌学检索词表》，特在此向李鸿恩老师及其同仁表示感谢！）

参考文献：

[1] 曾新红.《中国分类主题词表》的 OWL 表示及其语义深层揭示研究[J]. 情报学报，2005(2):151－160.

[2] ANSI/NISO Z39.19－2005. Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. Developed by National Information Standards Organization, approved July 25, 2005 by the American National Standards Institute.

[3] 曾新红等. 中文叙词表本体共建共享系统研究[J]. 情报学报，2008,27(3).

[4] 中华人民共和国国家标准. 汉语叙词表编制规则[S]. GB13190—91，北京：中国标准出版社，1991.

[5] Jena－A Semantic Web Framework for Java[EB/OL]. [2006－06－07]. http://jena.sourceforge.net/.

[6] Forgy C L. Rete: A Fast Algorithm for the Many Pattern/Many Object Pattern Match Problem[J]. *Artificial Intelligence*, 1982, 19(1):17－37.

[7] The RETE Algorithm[EB/OL]. [2006－06－10]. http://www.cis.temple.edu/~ingargio/cis587/readings/rete.html#2.

[8] 季羡林. 敦煌学大辞典[M]. 上海：上海辞书出版社，1998.

[9] SKOS Use Cases and Requirements: W3C Working Draft 16 May 2007[EB/OL]. [2007－07－08]. http://www.w3.org/TR/2007/WD－skos－ucr－20070516/.

（作者 E－mail:zengxh@szu.edu.cn）

中文叙词表本体一致性检测机制研究与实现

作者： 曾新红， 林伟明， 明仲， Zeng Xinhong， Lin Weiming， Ming Zhong
作者单位： 曾新红,林伟明,Zeng Xinhong,Lin Weiming(深圳大学图书馆,深圳,518060)， 明仲,Ming Zhong(深圳大学信息工程学院,深圳,518060)

刊名： 现代图书情报技术 PKU CSSCI
英文刊名： NEW TECHNOLOGY OF LIBRARY AND INFORMATION SERVICE
年，卷(期)： 2008，(5)
引用次数： 0次

参考文献(9条)

1. 曾新红 《中国分类主题词表》的OWL表示及其语义深层揭示研究[期刊论文]-情报学报 2005(02)

2. ANSI/NISO Z-39.19-2005.Guidelines for the Construction,Format,and Management of Monolingual Controlled Vocabularies 2005

3. 曾新红 中文叙词表本体共建共享系统研究 2008(3)

4. GB 13190-1991.汉语叙词表编制规则 1991

5. Jena-A Semantic Web Framework for Java 2006

6. Forgy C L Rete:A Fast Algorithm for the Many Pattern/Many Object Pattern Match Problem 1982(1)

7. The RETE Algorithm 2006

8. 季羡林 敦煌学大辞典 1998

9. SKOS Use Cases and Requirements:W3C Working Draft 16 May 2007 2007

相似文献(9条)

1. 期刊论文 曾新红.明仲.蒋颖.林伟明.胡振宁.张水英.Zeng Xinhong.Ming Zhong.Jiang Ying.Lin Weiming.Hu Zhenning.Zhang Shuiying 中文叙词表本体共建共享系统研究 -情报学报2008,27(3)

本文阐述了中文叙词表本体(OntoThesaurus,即基于中文叙词表建立的本体知识库)共建共享系统的设计思想和总体结构.描述了中文叙词表转换为OWL本体的扩展TBox定义,叙词表文本的ABox实例自动转换,OntoThesaurus的一致性检测机制;OntoThesaurus在图书情报界及语义Web界的广泛共享应用前景;在共享应用中采集标引员、领域专家和一般检索者知识实现本体共建和动态完善的完整过程.最后对我国叙词表编纂机构快速实现现有中文叙词表(主题词表)的网络化共建和共享服务提出了建议.

2. 学位论文 谷建军 基于叙词表的中医古籍文献领域本体建模方法研究 2006

1.前言随着20世纪90年代中医药文献数字化研究的开展，中医古籍文献数字化工作已经走过了几个阶段。从2000年国家中医药管理局设立的重点研究专项“中医药古代文献资源数字化关键问题研究”的起步阶段，到2001年国家科技部基础工作重大项目“中医药科技信息数据库建设”项目，再至2003年国家科技部医学科学数据共享服务系统“中医药学科学数据共享服务中心”建设项目，中医古籍文献数字化已成功研制出“中医本草文献数据库”、“中医方剂文献数据库”，在全国三十余家中医院校和和研究机构的参与下，成功构建了我国第一个中医古籍文献知识库，目前已收录了本草、方剂类古籍260余种，6000余万汉字，并于2003年实现了网络运行。 在数字化工作的研究中，导师柳长华教授提出的基于“知识元”的中医古籍计算机知识表示方法在知识库建设中取得了进展，基本形成了一套较成熟的建库技术。 以这种技术建立的数据库使知识的查询更加精确，避免了大量冗余信息的出现，使用户最大限度地摆脱了信息爆炸的困扰。但随之而来的另一个问题又出现在查询者面前，这就是所谓的“信息孤岛”现象。古籍数字化的功能不仅在于一般的信息查询，更重要的是古籍文献中的知识发现。普通的数据库难以达到知识挖掘的深层次要求，古籍数字化的目标是建设知识库。 2.知识库系统的原理从知识的使用角度来看，知识库是由知识和知识处理机构组成，知识库形成一个知识域，该知识域中除了事实、规则和概念之外还包含各种推理、归纳、演绎等知识处理方法。 知识库系统的核心组成部分是知识库和推理机构。知识库对知识进行存储和管理，推理机构是推理机使用知识库内的知识执行推理的机构。如果一个系统具有能用计算机所存贮的知识对输入的数据进行解释和推理，并有对其进行验证的功能，则该系统称为知识库系统。 知识库系统的实现涉及到两个关键问题：知识表示和知识推理。知识库的处理过程分为二个层面：先将知识由底层数据经过一系列加工，如分类、归纳、综合等处理过程而得到上层信息，称为知识表示。这种信息再经过解释、比较、推理得到我们所获取的知识，即知识推理的过程。 为了实现知识推理，一种基于本体的知识表示方法成为各个领域构建知识库推理系统的首选。 3.本体的概念、作用与分类本体(Ontology)起源于哲学领域，古希腊哲学家亚里士多德(Aristotle)定义Ontology为“对世界客观存在物的系统的描述，即存在论”。Ontology是客观存在的一个系统的解释或说明，它关心的是客观现实的抽象本质。Ontology这个哲学范畴，被人工智能界赋予了新的定义，从而被引入信息科学中。 目前普遍接受的本体定义为：共享概念模型的形式化规范说明。从内涵上来看，本体是领域(可以是特定领域的，也可以是更广的范围)内部不同主体(人、机器、软件系统等)之间进行交流(对话、互操作、共享等)的一种语义基础，即由本体提供一种明确定义。Ontology自身所要实现的目标，即：“在人类和应用系统之间实现共享和相互理解”。 Ontology能够将领域中的各种概念及概念之间的关系显示地、形式化地表达出来，从而将术语的语义表达出来，因而在语义查询方面发挥着重要作用。自W3C主席TimBermem-Lee在1998年首先提出了语义web的概念之后，Ontology正在成为人工智能和信息处理领域的研究热点之一。 本体强调相关领域的本质概念，同时强调这些概念之间的关联。本体论可以有效地表达知识和知识之间的关系，基于本体论的知识库系统可以建立有效的知识表达体系，揭示知识之间的内在关系。 本体技术主要在以下几个方面提高知识库系统的性能：可重用性、知识获取、查找智能性、可靠性、规范定义、任务解析、可维护性。 本体通常可分为以下几类：领域本体、通用本体、应用本体、表示本体。本文关注的是本体类型中的领域本体，主要讨论如何运用Ontology技术构建中医古籍领域本体。 4.本研究的意义、方法与创新点本文通过对本体的国内外研究与发展现状的考察，根据中医古籍数据库的实际情况，在知识推理层面提出了建设面向中医古籍数据库应用的中医古籍文献领域本体的设想。参考国内外领域本体的建设方法，论述了利用叙词表建设领域本体的优势，提出了基于叙词表的适合中医古籍数据库应用的中医古籍文献领域本体建设方法。最后通过一个实例阐述了中医古籍文献领域本体的具体建设方法，为中医古籍数据库的进一步建设提供了理论与实践的双重参考。 研究意义：中医古籍知识库建设的要求；中医古籍知识深入整理研究的要求；便于网络中医古籍文献资源的统一管理。 研究方法：文献调研法、概念分析法、本体构建法。创新点：在中医古籍文献数字化领域提出建立本体系统的设想；分析了适合中医古籍文献数据库的本体表

示语言和编辑工具；提出中医古籍文献领域本体的建设目标；设计了中医古籍文献领域本体的建设方法；建立了一个以“病证”概念为核心的中医古籍文献领域本体模型。 5.本体的国内外研究现状国外主要研究现状：①理论深化研究；②信息系统中的应用；③本体作为一种能在知识层提供知识共享和复用的工具在语义网中的应用。 国外较为知名的本体知识系统：WordNet、FrameNet、GUM、SENSUS、OntoSeek、Cyc、HowNet和SUMO等。国内主要研究现状：我国本体的研究尚处于起步阶段，一个是对W3C发布的关于本体的外文资料的翻译，一个是主要为面向应用的研究，无论是理论还是实际应用都相对落后于国外。 面向中医药领域的研究主要有：浙江大学网格计算实验室开发的基于语义的中医药信息本体虚拟组织模型——DartGrid服务栈；北京中医药大学和中国科学院计算机研究所开发的基于本体的中医专家临床病案知识库。 6.领域本体的构建20世纪50年代叙词表得到了很大发展，成为主题检索的主要语言，各国拥有的叙词表数以千计，并涵盖了各个领域。从一定意义上讲，叙词表可以说是一种轻量级本体(Light-weightOntology)。基于叙词表构建领域本体有诸多的优越性，目前人工智能界普遍推荐利用叙词表构建领域本体。 中医古籍文献叙词表与本体的关系：中医古籍文献叙词表表示的是中医古籍文献中包含的概念，概念来自于古籍内容与古籍本身，是对中医古籍文献的客观反映。叙词表表示的是树状结构，这种树状结构反映了古籍文献内部的自然构成方式。叙词表的结构是可见的、清晰的，可称为显性结构。领域本体继承了叙词表的树状结构特征。本体更重在表示一种概念之间的隐含关系，这种关系是模糊的，不明显的，可以称为隐性结构。相对来说，本体的反映更微细，更深入，为文献中的知识关联提供了可实现的途径。叙词表或本体是对体现古籍内涵的概念的集合。 领域本体的建模元语：(概念)类、属性、函数、公理、实例。 建模语言：选用OWL语言。本语言的优势在于：基底层语法符合XML标准格式；为W3C推荐的标准本体编辑语言，便于与数据库之间的数据交换；支持多种语言输入，并支持中文；网络中有免费教学手册，便于下载学习。 编辑工具：选用Protégé-2000。其优势在于：界面友好，具有图形化的用户界面；版本更新速度快，目前已发布了3.1.1版；支持多种语言格式，支持中文编辑；本体文档可以不倚赖于本体编辑器进行代码修改，方便与数据库的连接；网络开放资源；是W3C推荐的本体编辑器；是基于XML的本体标记语言，多种存储格式，可以适应不同需要。 构建方法：选用斯坦弗大学医学院开发的七步法。7.中医古籍文献领域本体模型(病证模型)的构建元数据(Metadata)就是数据之数据，或描述原始数据的独立数据。元数据是针对网络信息标引发展起来的，它以Web页作背景，通过元数据将Web信息组织起来，构成基于元数据的有序信息系统，为网络信息资源的组织提供了重要手段。其主要学术意义和应用价值在于信息处理。 根据中医文献数字化研究室的最新研究，中医药古籍元数据包括三类概念：一是表达古籍外部特征的元数据，称为书目元数据；二是表达古籍内部篇、卷、章、节层次特征的元数据，称为书体结构元数据；三是表达古籍知识单元内容的元数据，称为语义元数据。本领域本体模型以“语义元数据”为核心概念集，以“病证”语义元数据及其包涵的概念为中心建立本体模型。

有关病证与其他概念间的关系主要有二类：等级关系，包括上下位关系和实例关系；非等级关系，包括同义关系、交叉关系、排斥关系等。以《诸病源候论》“风痉候”为例，为本体添加类和实例：“风痉候”条文：“风痉者，口噤不开，背强而直，如发痫之状。其重者，耳中策策痛；卒然身体痉直者，死也。由风邪伤于太阳经，复遇寒湿，则发痉也。诊其脉，策策如弦，直上下者，风痉脉也。” “风痉候”的概念等级链为：病证——风病——风痉。条文中与与本概念相关的其他概念有：证候表现、预后、病因、病位、脉象。添加到本体中，如图所示： 8.讨论中医古籍文献领域概念十分丰富，概念间关系错综复杂，难以在短时间内完成本体系统的建设，应根据实际需要分阶段完成。本文将中医古籍文献领域本体的研究目标分为二个阶段： 长期目标：建立相对完整的中医古籍文献领域本体系统平台。建立本体的中英文对照词表，便于与世界接轨。 短期目标：根据数据库建设的需要，分别以本草、方剂、病证为中心概念，开始本体系统的建设。

3.期刊论文 曾新红.Zeng Xinhong 中文叙词表本体——叙词表与本体的融合 -现代图书情报技术2009(1)

从网络信息社会对知识组织系统的需求、来自信息科学界和其他相关各界的应对发展现状等方面,详细阐述实现中文叙词表的形式化表示和网络应用的重要性和迫切性.对叙词表和本体的概念进行深入的比较研究,论证将他们合二为一的可行性.阐述直接采用OWL(而不用SKOS)表示中文叙词表本体(OntdThesaurus)的原因,并列出具体的类定义和属性定义.中文叙词表本体共建共享系统OTCSS的多项功能和若干原型系统的实现,证明这些定义的科学性、可行性和通用性.

4.期刊论文 曾新红.林伟明.明仲.Zeng Xinhong.Lin Weiming.Ming Zhong 中文叙词表本体的检索实现及其术语学服务研究 -现代图书情报技术2008(2)

在简单介绍中文叙词表本体共建共享系统OTCSS项目背景的基础上,阐述实现中文叙词表本体网络术语学服务(OntoThesaums -TS)的意义.详细描述OntoThesaums的检索实现方法,以及其术语学服务应用场景典型案例,并对OntoThesaurus的术语学服务提出进一步研究计划.

5.学位论文 付佳佳 基于叙词表的领域本体建模研究 2006

众所周知，叙词表是一种为解决信息的主题排序而创造的人工语言，它的本质是对自然语言中的词汇进行选择、规范、并揭示其间相关关系，由此形成受控词汇的集合，它的出现主要是为了解决大量的文献如何被方便科学检索的问题。然而，WWW是当今主要的网络信息的集散地，不仅汇聚了海量的信息，而且信息数量正在以指数级的速度增长。随着数据量的激增，WWW上大量分布的无结构和半结构化数据日益加剧信息检索的困难，因此，如何组织海量的数字信息，并为用户提供精确高效的网络检索服务成为重要而迫切的研究课题，这引起了人们对传统知识组织工具如叙词表、分类表等在网络环境中适应性的争论。尽管叙词表和分类法等传统知识工具已开始在网上发展，但是对机器语言来说，其互操作性和表达性仍比较差，为此人们提出了本体这种能在语义和知识层次上描述信息系统的概念模型建模工具。领域本体构建的重要意义主要体现在： 首先，领域本体的目标是捕获相关领域的知识，确定该领域内共同认可的词汇，并从不同层次的形式化模式上给出这些词汇之间相互关系的明确定义。从而实现人们对同一客观事物的共识，形成一个统一的认识事物的标准。即为人类认识活动构建顶层概念框架。 其次，本体更加突出知识共享的功能，尽管二者都对概念间等级关系、相关关系进行了揭示，但本体更着眼于给出人类事物认识的知识(或领域知识)总框架，因为在本体的一个实体中每个概念都有其属性信息、实例信息，而这些在词表系列中则少有展示，很多已经涉及到专业词典中的知识，因此说一个本体是一个人类知识(或领域知识)体系的汇总毫不夸张。 最后，本体的出现还是为了设计一种机器可以理解的语言。通过本体可以克服计算机系统之间的语义鸿沟，实现某个领域内不同主体(人、机器、软件系统等)之间的对话、互操作、知识共享等目的，于是它被认为是一种共享的概念模型的形式化的规范说明。其中形式化就是指应该是机器可读(可理解、可操作)的意思，而这也成为了在计算机网络环境下应用研究的主题之一。 领域本体的构建体现了目前的趋势，但是原本属于本邻域的叙词表是丢弃还是融合?这是本文探讨的问题。笔者认为，由于叙词表和领域本体之间有许多的相同和不同之处，使得基于叙词表来构建领域本体具有一定的优越性。由于某学科领域的叙词表包括本学科领域中相对比较完整的术语(叙词)，因此这些术语(叙词)可以为本领域本体中的概念的创建提供指导；另外，叙词表中的限义词、涵义注释、等级关系、词间关系，为领域本体中概念的属性、实例以及关系的创建可以提供线索和指导，这些指导将为领域本体的创建者们节省大量的时间和精力。基于叙词表构建的领域本体至少在本领域的概念方面应该是比较完整的。叙词表可以说是图书馆情报界为信息检索提供的知识财富，其作用和原理与本体有异曲同工之妙。如果能利用现存的叙词表，将其转换为相应的领域本体，必将使领域本体的建立事半功倍。 本文在第一章中，研究了本体在改善对知识管理方面的作用，论述了建立领域本体这一课题的意义，阐述了本文的研究内容和本文的章节安排。在第二章中，系统地研究了本体的理论，从本体的定义、分类、描述语言和建模工具等方面进行了论述。而在第三章，研究了叙词表的概念、应用现状，并分析了叙词表在表达语义方面的局限性和本体在此方面的优势。为了提高论述的有效性，本文还以具体的例子来说明了这一点。在第四章，根据前文的论述，总结并分析了叙词表和领域本体的区别与联系，论述了基于叙词表建立领域本体的可行性和优越性。第五章，本文又研究了当前本体建模的主要的方法，并在总结这些方法的特点的基础上，提出了基于叙词表建立领域本体的方法。在第六章，本文通过对食品安全领域本体的建模这一实例，详细地说明了这一方法。在这个实例中，笔者自行开发了一个由叙词表的词间关系向领域本体的概念间关系转化的系统，从而实现了基于叙词表建立领域本体的关键一步，这也是本文的创新之处。第七章，文章对全文作了一个总结，提出了本文的不足之处以及对未来工作的展望。 本文采取了文献调查、案例佐证、技术对比等方法，从理论和实践的角度研究了基于叙词表的领域本体的建模。但是，由于国内对于本体的开发方法以及如何构建领域本体研究的较少，对基于现有的叙词表构建领域本体的研究，也还处于起步、探索阶段。同时限于个人的能力和水平，笔者仅对本体及叙词表的理论，基于叙词表构建领域本体的可能性及方法进行了相当粗浅的研究；另外，由于任何一个领域本体的构建都是相当复杂的，而且需要该领域的专家的参与，同时还要耗费大量的人力和物力，不是一个人在短期内就能完成的，因此，笔者开发的系统还没有广泛地得到验证，所构建的领域本体模型也比较简单，一部分功能还没有完全实现，还需要进一步的完善。这些都是笔者将来要做的工作。

6.学位论文 鲜国建 农业科学叙词表向农业本体转化系统的研究与实现 2008

在当今信息时代和知识经济社会，信息资源已成为重要的战略资源，在国家科技进步与创新、经济和社会可持续发展过程中发挥越来越重要的作用。现代信息技术和通信技术为信息的收集、加工、存储、传输和利用提供了强有力的技术保障，信息资源呈指数级增长。大量的信息给人们的工作、生活和学习提供丰富的信息资源的同时，又使人们淹没在信息的海洋之中。如何组织、管理和维护海量的信息资源并为人们提供高效优质的信息服务成为一项重要而迫切的任务。本体(ontology)作为一种能在语义和知识层次上描述信息系统的概念模型建模工具，为解决这一问题提供了新的途径，已受到

国内外研究人员的广泛关注，成了研究的热点，而本体构建也是其中一个重要的研究方向。　　本文对本体和叙词表的相关知识进行了详细论述，分析得出了叙词表向本体转化的必要性和可行性，并使用当前最新的本体描述语言—网络本体语言(WebOntologyLanguage，简称OWL)，成功地将《农业科学叙词表》(以下简称《农表》)中的叙词(包括正式叙词及非正式叙词)及词间关系进行了表示和描述。在此基础上，设计和实现了一个转化系统，能够自动批量地将词表中的知识结构和语义关系转化到农业本体中。基于叙词表来构建领域本体，不仅为构建领域本体提供了一种较好的方法，也可以加快本体的构建进程，还能提高本体的科学性、规范性和权威性。　　本文还在本体应用方面进行了探索，基于转化得到的农业本体构建了一个智能检索原型系统，在智能导航、自动扩大检索范围和跨语言检索等方面都进行了初步尝试。实验结果表明，该系统能提供较友好的导航功能，检全率也有一定的提高，还可以实现简单的跨语言检索。如果能进一步丰富和完善这些功能，将能大大提高传统检索系统的性能，也将会有广阔的应用前景和实际的使用价值。

7.期刊论文 李娜.任瑞娟 叙词表、分类法与分布式本体 -现代情报2007,27(12)

本文分析叙词表、分类法与分布式本体概念的内涵与外延及各自的属性,探讨了三者相互关系,在此基础上提出了建立基于叙词表、分类法与分布式本体模型的设想.这种分布式本体是在语义和知识层次上描述信息系统的概念模型建模工具.通过对这种分布式本体的机理与实现方法的分析与总结得出结论:基于叙词表、分类法构建的分布式本体是在分布异构的网络环境下探索知识发现、知识组织、知识检索、知识服务的有效途径,是智能网络服务的必然归宿.

8.会议论文 曾新红.林伟明 中文叙词表本体共建共享系统OTOSS的设计与实现 2007

阐述了中文叙词表本体(OntoThesaurus,即基于中文叙词表建立的本体知识库)共建共享系统OTCSS的设计与实现方法,并对我国叙词表编纂机构利用本系统快速实现现有中文叙词表(主题词表)的本体转换和网络化共建共享提出了建议。

9.期刊论文 Choi Suk-Doo.王一丁 利用叙词表开发本体 -数字图书馆论坛2007(5)

文章提出了一种构建大规模韩语叙词表的方法,这种叙词表可用于在各种不同领域内提高检索性能.目前它主要用于标引以及检索过程,新的词汇也正源源不断地添加进来.随着韩语中对于检索性能的新需求的不断增加,开发一个大规模的本体系统应当是必要的,因而一个正在进行的项目的目标就是把现有叙词表转变为一个本体系统.文章将描述叙词表是如何构建的,并指出如何将其演变为一个本体系统的基础.