

知识本体的检索机制研究

赵 冬^{1,2}, 牛 强¹, 刘晓明¹

(1 中国矿业大学 计算机科学与技术学院, 江苏 徐州 221116;

2 山西潞安集团技术中心, 山西 长治 046204)

摘 要: 针对知识本体的相关应用, 结合知识检索的相关内容, 分析了知识本体检索中存在的问题, 构建了知识本体的检索机制, 给出了基于本体的知识检索流程, 并重点对检索中查询构造器、查询引擎及结果定制模块进行了详细的分析, 给出了各个模块的具体结构及实现方法. 最后结合提升机故障知识本体, 通过系统的实际应用说明该方法有利于增加提升机故障知识的复用性和通用性, 同时也验证了所提出的方法的有效性和正确性.

关键词: 知识本体; 知识检索; 提升机故障

中图分类号: TP311

文献标识码: A

文章编号: 1000-7180(2011)10-0129-04

Research on Knowledge Ontology Retrieval Mechanism

ZHAO Dong^{1,2}, NIU Qiang¹, LIU Xiao-ming¹

(1 School of Computer Science & Technology, China University of Mining and Technology, Xuzhou 221116, China;

2 The Technique Center of Shanxi Luan Group Co., Ltd, Changzhi 046204, China)

Abstract: For the knowledge ontology related application, combined with the relevant content of knowledge retrieval, the problems of knowledge retrieval is analyzed. The retrieval mechanism for the hoist fault knowledge ontology is constructed; the retrieval processes of ontology-based hoist fault knowledge is given; a detailed analysis of the query constructor, query engine and the results custom modules is focused; the specific structure and realization of each module is given. Finally, combined with hoist fault knowledge ontology, the implementation and application of system show this method help to increase the reusability and generality of the hoist fault knowledge; the system can provide users with comprehensive and accurate domain knowledge and also verify the method is effective and correct proposed in this thesis.

Key words: knowledge ontology; knowledge retrieval; hoist fault

1 引言

从知识表示角度考虑, 本体作为一种能在语义知识层次上描述信息系统的概念模型建模工具, 能很好地描述一个具体或是抽象领域中包含的知识的定义、表示和组织. 国内外已经有许多研究者将本体引入知识工程领域^[1], 如 WordNet 项目^[2], 它是一种基于认知语言学的英语词典, WordNet 在信息检索和知识工程的应用, 为这两个领域开启了新的愿景; SWAN (语义网应用于神经医学) 项目^[3]; SWAN (Semantic Web Applications in Neuromedi-

cine) 是一个跨学科项目, 通过语义网技术启用社会的能源和自组织能力, 为阿尔茨海默病 (AD) 研究界制定一个有效的专门知识基地; 知网项目^[4], 知网是国内第一个电子知识系统, 是一部比较详尽的语义知识词典, 自问世以来受到了很多专家学者的普遍重视.

在本体应用于知识工程中, 自然知识本体的检索也逐渐成为学者的研究热点. 不同领域的知识本体表示, 因为都是基于本体, 所以对知识本体的检索机制研究有助于提高知识的应用和检索效率. 文献 [5] 中从知识检索的角度阐述了知识检索的思想, 并

对知识检索中的语义检索进行了描述;文献[6]中对本体的基本研究问题进行了概述,其中就涉及到基于本体的语义检索问题.通过以上分析,本文在对知识本体检索系统结构研究的基础上,详细的阐述检索机制中涉及到的各个模块.主要内容包括提升机故障知识本体检索的系统结构,以及查询过程中查询构造器的设计、查询引擎的设计和查询结果组织等问题.

2 知识本体检索的系统结构

较好的知识检索系统能充分的体现实际的需求,从而很好的帮助相关领域的工作人员提高工作效率.本文根据矿井提升机系统在煤矿中的实际工作情况,结合相关文献中的有关语义检索的知识^[6-7],给出提升机故障知识本体检索的系统结构如图1所示.

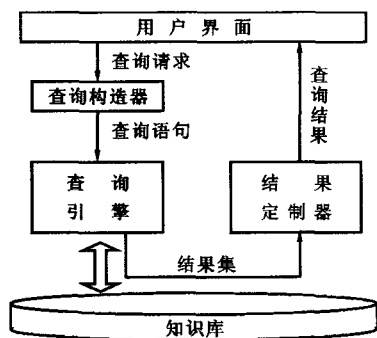


图1 提升机故障知识本体检索的系统结构

提升机故障知识本体检索的处理流程如下:

用户通过用户界面输入查询请求,查询构造器负责对用户的查询请求进行转换,然后将转换成功的查询语句提交给查询引擎;查询引擎负责访问本体知识库,建立知识模型,解析查询语句和执行查询,获取查询结果并将查询结果提交给结果定制器;结果定制器负责将查询结果进行格式转换,然后呈现给用户.

从图1可以看出,矿井提升机故障知识检索模型的功能模块主要由用户界面、查询构造器、查询引擎、结果定制和知识库五部分组成.本文认为基于本体的知识检索主要机制包括:查询构造器、查询引擎和结果定制三部分.

3 查询构造器设计

图2给出的是查询构造器的结构图.

从图2可以看出,系统首先判别用户选择的是

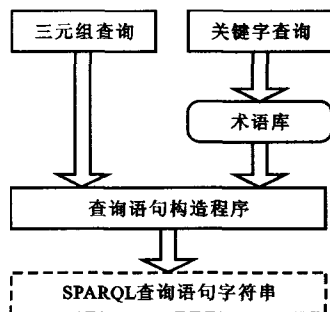


图2 查询构造器结构图

三元组查询还是关键字查询.如果用户选择的是关键字查询,那么还应通过查询扩展程序访问术语库(包括概念、属性和实例的词语库,术语库的存储通过关系数据库来实现)来进行匹配,通过这种方式来得到构造查询模式所需要的字符串.然后再转交给查询语句构造程序,通过查询语句构造程序来构造查询语句字符串.如果用户选用的是三元组查询,因为三元组中的查询词是通过解析本体得到的,所以可以直接通过查询语句构造程序来构造查询语句字符串.

本文选用的是 SPARQL 查询语言. SPARQL 是 W3C 的 RDF 数据工作组设计的一种查询语言和协议,是一种从 RDF 图中获取信息的查询语言.它可以用来对 URIs、空节点和无格式类型文本提取信息、RDF 子图以及根据查询图中信息构造新 RDF 图.它的主要功能包括:提炼查询结果、可选匹配、值约束条件、替换匹配、以及直接回答 YES/NO 等.本文主要讨论的是 SELECT-FROM-WHERE 语句来对本体知识库进行查询.本文的查询语句构造处理流程如图3所示.

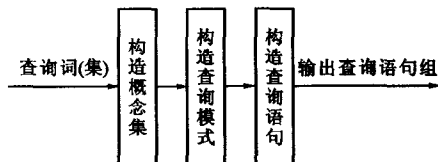


图3 查询语句构造处理流程

从图3可以看出在查询语句的构造中主要包括三个处理阶段:构造概念集、构造查询模式和构造查询语句.

概念集主要是将用户输入的查询词(集),经过一系列的转换得到各自对应的本体中的概念^[8].根据用户输入的查询词在本体中对应的类别,本文将概念集主要分为四类,分别是:CCSet、PCSet、ICSet及PICSet.其中CCSet存放的是根据查询词对应本体中的类扩展得到的概念集;PCSet存放的是根据

查询词对应本体中的属性扩展得来的概念集;ICSet存放的是根据查询词对应本体中的实例扩展得来的概念集;PICSet 比较特殊,它是在对 PCSet 和 ICSet 进行概念合并时产生的新的概念集。

SPARQL 检索语言主要通过图形模式的匹配来实现查询功能。最简单的图形模式是三元组模式 $\langle \text{Subject}, \text{Predicate}, \text{Object} \rangle$, 并允许变量出现在三元组的任意一个位置上。因此,查询模式构造主要是解决概念、实例和属性如何转换成 SPARQL 查询模式的问题。

用户在查询时并不关注本体的构造,而是要查出领域知识部分即本体的实例数据。所以查询模式的构造也着重体现在对实例的查询上,这就引发了对概念集中重复概念的合并问题,在查询语句构造阶段称为概念集预处理。

通过查询模式构造规则对概念集的处理,可以得到 SPARQL 查询模式。利用这些查询模式,就可以完成对查询语句的构造。查询语句的构造示意图如图 4 所示。

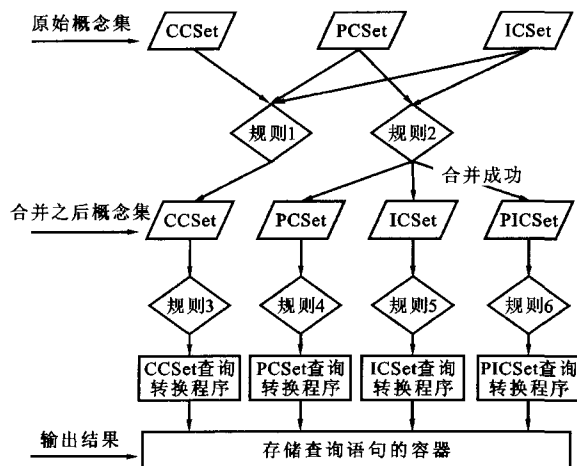


图4 查询语句构造示意图

从图4可以看出,查询构造器为每一种概念集都单独设计了查询转换程序,这样系统可以根据不同的概念集合进行简单而合理的转换。结合 SPARQL 查询语言的规则,通过查询模式构造规则的处理,生成的模式就可以直接作为 SPARQL 中的 WHERE 后面的查询条件,而 SELECT 后面的查询变量就对应模式中的变量。至此,通过以上的方法就可以很好的完成从用户查询词到 SPARQL 查询语句的转换。

4 查询引擎设计

本体的推理功能是它区别于其他知识组织体系

的优势所在。本体推理作为一类应用,可以映射到规则推理引擎上进行推理^[9]。利用建立的本体对领域知识进行标注时,描述知识的各种词汇映射到本体的概念中,通过本体概念间的各种关系,词汇间建立语义关系,从而形成网状的知识库,为进一步的推理操作打下基础。

本体查询引擎具有两种典型的体系结构,一种是传统的独立式查询引擎,另一种则是查询问答系统。而本文的查询系统主要是满足知识库检索的需要,因此本文采用了传统的独立式查询引擎设计。图5是查询引擎的结构图。

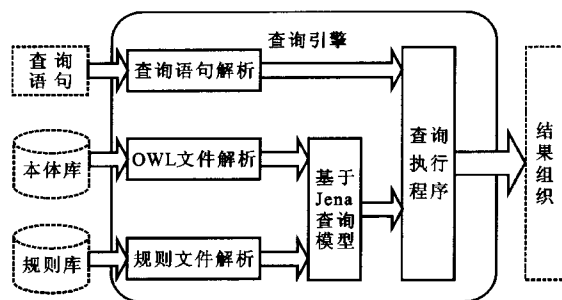


图5 查询引擎结构图

用户通过查询引擎实现的查询步骤如下:查询引擎先通过 Jena 工具包对本体库和规则库进行解析,然后建立基于 Jena 的查询模型。查询引擎接着利用从查询构造器传递过来的查询语句,通过 Jena 提供的查询工具进行本体查询。

本体模型是领域本体知识库在内存中的形式,本体知识库以 OWL/XML 的形式存储在磁盘中,操作时需要将其读入内存生成结构模型,利用模型的各种方法完成修改、检索、推理等操作。其中,领域规则解析程序负责调用和解析规则库文件并实现基于它的一般规则推理,其作用就相当于传统的规则推理引擎。这样就可以使得系统能处理更复杂的提升机故障领域知识的能力;OWL 文件解析模块负责对本体资源图实现基于规则的语义扩展,这样就可以增加用户在推理方面的能动性。Jena 查询模型根据事实库中的已有事实,通过解析后的规则推理出新的事实,并将新的事实经过格式转换加入到 OWL 知识库中,以达到对 OWL 知识库内容的更新与扩充。

5 查询结果组织

查询结果定制器主要负责将查询引擎的查询结果经过处理之后反馈给用户。一般情况下,不论是使用 SPARQL 查询还是用推理机进行推理的结果都

是使用带有命名空间的文本格式,如果直接呈现给用户而对结果不加处理,那么用户很难读懂其中的意思.所以必须经过转化才能以比较清晰而且简明的二维表格式呈现给用户.

大多数的 SPARQL 处理器将包括一个定制的 API,允许直接操纵结果集,以适用实际中的应用.但是,如果要以一种标准化的方式序列化 SPARQL 结果集,可以使用 SPARQL 查询结果的 XML 格式输出^[10].通过以上分析,本文设计出如图 6 所示的查询结果定制的处理流程.

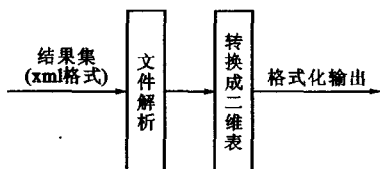


图 6 结果定制处理流程

由图 6 可以清楚地看出结果定制的处理流程:文件解析程序先对输入的 xml 格式字符进行解析,然后根据转换关系将 xml 格式文件以二维表的格式输出.

6 应用分析

本文以提升机故障知识本体为例,来介绍知识本体检索的系统应用问题.系统研究的目标是开发一个可以为煤矿生产企业的一般技术人员提供提升机故障知识检索的系统. Java 作为一种与平台无关的编程语言,近年来被开发者广泛的使用.另外开发过程中用到的开发包 Jena 和 Pellet 也是用 Java 语言开发的工具包,因此选择 Java 作为 HFKRS 系统的开发语言.开发平台及环境如下:

- > 开发平台: Window XP
- > 开发语言: Java
- > Web 服务器: Tomcat 5.0
- > 数据库服务器: Sql Server 2000
- > 第三方包工具: JDK 1.5.0 版, Jena 2.5, Pellet 2.0
- > 开发工具: MyEclipse 6.0

限于篇幅的原因系统的具体实现在此就不在赘述.结合本体的设计思想,给出知识检索中的三元组模式检索界面图(图 7),此功能是系统所有用户的权限,通过该功能用户可以通过概念、属性和实例中的一种或者几种组合起来的方式来检索知识库,并对检索结果进行结果定制.

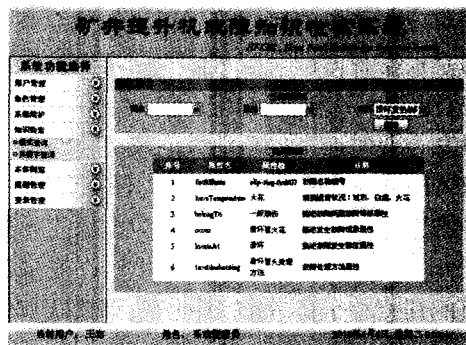


图 7 三元组模式检索界面

7 结束语

构建好的领域知识库,必须有一套合理而且完善的检索机制来构建整个知识检索系统,这样才能体现针对矿井提升机故障领域知识表示的优越性.本文从基于本体的知识检索系统结构出发,分析了构建一个理想的基于本体知识库的知识检索系统要解决的问题并给出了检索系统的处理流程;接着给出了基于三元组模式和关键字的查询构造器设计;然后设计并构建了符合本文思想的查询引擎;最后为了更好的把查询结果呈现给用户,对查询结果进行了定制.通过以上的的工作给矿井提升机故障知识模型的检索研究提供了有效的支援和保障.通过原型系统的实现和应用,说明该方法有利于增加提升机故障知识的复用性和通用性,能够为用户提供全面准确的提升机故障领域知识,同时也验证了本文所提出的方法的有效性和正确性.

参考文献:

- [1] 余倩. 近年来领域本体的应用新进展[J]. 图书馆建设, 2008(8): 95-99.
- [2] Princeton. WordNet: a lexical database for the English language[EB/OL]. [2011-04-16]. <http://wordnet.princeton.edu>.
- [3] Paolo Ciccarese, Wu Elizabeth, Gwen Wong et al. The SWAN biomedical discourse ontology[J]. Journal of Biomedical Informatics, 2008(41): 739-751.
- [4] 董振东, 董强. 知网: HowNet Knowledge Database[EB/OL]. [2011-04-16]. <http://wordnet.princeton.edu>.
- [5] 马文峰, 杜小勇. 知识检索研究[J]. 情报理论与实践, 2006(2): 157-161.
- [6] 丁晨春, 顾德访. Jena 在实现基于 Ontology 的语义检索中的应用研究[J]. 现代图书情报技术, 2005(10): 5-9.
- [7] 许浩. 基于本体论的语义检索技术的研究与实现[D].

(下转第 136 页)

具有以下两个重要特性:

(1) 最小角最大: 在所有剖分方法产生的三角形中, Delaunay 三角剖分产生的三角形, 其最小内角之和最大。

(2) 空外接圆: Delaunay 三角剖分产生的任意一个三角形的外接圆, 都包含且仅包含形成该三角形的三个点, 而不包括任意其他点。

因此, Delaunay 三角剖分形成的三角网格是唯一的最优的网格。

现在有很多种算法计算 Delaunay 三角剖分, 其中一些很有效, 但内部描述却很难描述。其中一种较为简单且容易实现的算法步骤如下^[6]:

(1) 添加外部三角形, 并从它的一个顶点处开始。

(2) 加入一个内部点, 在三角形的外接圆内搜寻该点, 去掉包含该点的三角剖分。

(3) 重新构造三角图, 在刚刚去掉的三角形外接圆内包括新的点。

(4) 返回(2), 直到没有新的点加入。

本文实验采用 Delaunay 三角剖分后所得的三维重建效果如图 6 所示。

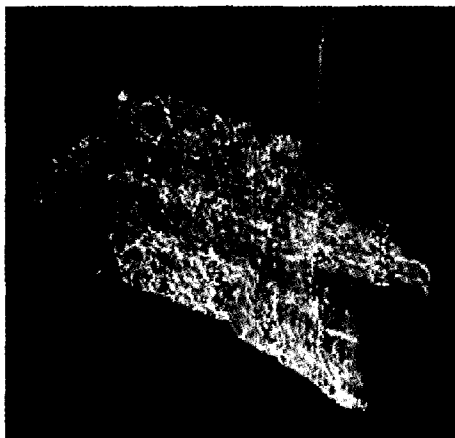


图 6 三维重建效果图

由图 6 可以看出, 有很多飘逸点影响了模型的恢复效果, 这是因为在角点匹配过程中, 匹配点中存

在一些误配点, 这主要是由于匹配图像中相似特征的干扰造成的, 这些可以通过增大匹配窗口和提高图像精度方面来解决。

6 结束语

本文出于快速建立煤矿巷道三维模型的考虑, 初步探讨了使用基于数字图像的三维场景重建技术的可能性, 并利用 OpenCV2.1 视觉库开发了一套基于巷道图像重建三维模型的系统, 并取得了初步效果。三维重建是计算机视觉研究的热点, 它直接模拟人的双目观察景物的方法, 获取景物的立体信息, 对它的研究, 无论是在仿生角度还是工程应用角度都具有重要意义。而本文提出把三维重建技术应用到煤矿巷道的建模上, 有助于巷道三维模型的快速建立, 进而为虚拟矿山的漫游奠定了基础。

参考文献:

- [1] 薛耀红. 基于二维图像的人脸三维数据获取[D]. 长春: 吉林大学, 2007.
- [2] Zhang Zhengyou. A flexible new technique for camera calibration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1330-1334
- [3] 陈乐, 吕文阁, 丁少华. 角点检测技术研究进展[J]. 自动化技术与应用, 2005, 24(5): 1-8
- [4] 孙卜郊, 周东华. 基于 NCC 的快速匹配算法[J]. 传感器与微系统, 2007, 26(9): 104-106.
- [5] 万琳. 基于三角网格的图像表示方法研究[D]. 武汉: 华中科技大学, 2009.
- [6] Gary Bradski, Adrian Kaehler. 学习 OpenCV (中文版)[M]. 于仕琪, 刘瑞楨, 译. 北京: 清华大学出版社, 2009.

作者简介:

山 维 男, (1986-), 硕士研究生. 研究方向为计算机视觉.

刘厚泉 男, (1963-), 博士, 教授. 研究方向为虚拟现实、计算机视觉.

(上接第 132 页)

南京: 南京理工大学, 2007.

[8] 聂卉. 基于本体的查询扩展与规范[J]. 现代图书情报技术, 2007(3): 35-38.

[9] 何琳, 曹玲. 农业古籍本体的构建及其检索机制研究[J]. 现代图书情报技术, 2006(12): 37-40.

[10] 何琳, 曹玲. 农业古籍本体的构建及其检索机制研究[J]. 现代图书情报技术, 2006(12): 37-40.

作者简介:

赵 冬 男, (1980-), 硕士生. 研究方向为知识本体研究.

牛 强 男, (1974-), 博士, 副教授. 研究方向为数据挖掘与知识获取研究.

刘晓明 男, (1987-), 硕士生. 研究方向为时间序列研究.