

提升机故障知识本体检索机制研究

许晓林^{1,2}, 牛 强¹, 林 伟¹

(1 中国矿业大学 计算机科学与技术学院, 江苏 徐州 221116;

2 龙煤集团鹤岗分公司峻德煤矿, 黑龙江 鹤岗 154101)

摘 要: 针对知识本体检索中存在的问题, 构建了提升机故障知识本体的检索机制, 给出了基于本体的提升机故障知识检索流程, 并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, 同时也验证了本文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(2012)03-0064-05

Research on Knowledge Ontology Retrieval Mechanism of Mine Hoist

XU Xiao-lin, NIU Qiang, LIN Wei

(1 School of Computer Science & Technology, China University of Mining and Technology,
Xuzhou 221116, China; 2 Heilongjiang Longmay Mining Group Co. Ltd., Hegang 154101, China)

Abstract: To the problems of knowledge retrieval with related application, The retrieval mechanism for the hoist fault knowledge ontology is constructed, the retrieval processes of ontology-based hoist fault knowledge is given, including query constructor, query engine and the results custom modules. 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, 能很好的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. 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[1]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[2]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.

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^[3-4],根据以上问题的讨论,给出提升机故障知识本体检索的系统结构如图1所示。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。其中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。

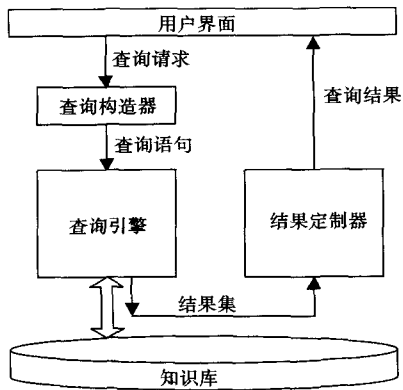


图1 提升机故障知识本体检索的系统结构

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。

3 查询构造器设计

3.1 查询构造器的结构

通过相关的研究发现,针对用户的一个具体查询,如果这个查询的内容相对于现有的知识库而言,知识库比较完善的话,便能得到很好的检索结果。反之,相对于这个用户查询,而知识库中不存在或存在的相应知识不够的话,其语义检索性能便比不上基于关键词的检索性能。这时,考虑将传统的基于关键词的检索策略整合进来,从而克服了知识库不充分时的不足,在整体上提升了检索性能。

结合本文的实际情况,系统的用户检索接口上采用三元组模式查询和关键字查询相结合的方式^[5]。通过模式查询用户可以清楚地表达自己的查询语义,系统通过自动获得概念、属性和实例的相关词汇作为检索词汇供用户选择,从而既提高了查准率又方便了用户。严格限定的模式查询在定义良好的知识库上能很好的保证系统查准率,但是由于模

式查询对提升机故障领域的本体库的要求很高,所以这样会使得系统的查全率得不到保证。将关键字查询作为三元组模式查询的补充,通过这样的方式来保证系统的查全率。在查询语言的选择上,本文选用的是 SPARQL 查询语言。图2给出的是查询构造器的结构图。

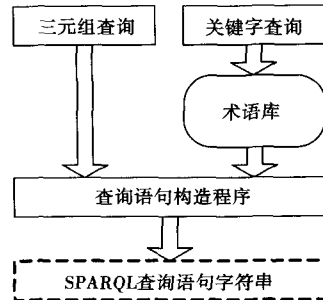


图2 查询构造器结构图

从图2可以看出,系统首先判别用户选择的是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。

3.2 查询语句的构造

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。本文主要采用 SELECT-FROM-WHERE 语句来对本体知识库进行查询,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。

(1) 概念集构造

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^[6]。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。

(2) 查询模式构造规则

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。本文设计出并遵循以下概念合并规则:

规则 1 若概念集 CCSet 某一概念与概念集 PCSet 中某一属性所对应的定义域概念、值域概念一致或者是其子概念时, 从概念集 CCSet 中删除该概念; 若概念集 CCSet 某一概念与概念集 ICSet 中某一实例所属概念一致或者是其子概念时, 从概念集 CCSet 删除该概念。

规则 2 若概念集 ICSet 中某一实例所属概念与概念集 PCSet 中某一属性所对应的定义域概念、值域概念一致或者是其子概念时, 将 ICSet 集中该实例与 PCSet 中该属性及所对应的概念存入新的概念集 PICSet 中。

在将概念集转换成查询模式中的组成部分时, 利用预处理之后的各个概念集, 本文设计出并遵循以下查询模式转换规则:

规则 3 若概念属于 CCSet 集, 约定将查询该概念的所有实例。

例如: 有概念“Class: 滑环故障”, 那么依据规则应被转换成下面的形式:

$\langle ? x, \text{rdf:type}, \text{Class: 滑环故障} \rangle$

规则 4 若概念属于 PCSet 集, 约定找出该属性的定义域和值域, 再查询与该属性相关的对应定义域和值域实例信息。

例如: 有属性“Property: troubleshooting”, 该属性在本体中定义域是“Class: 故障描述”, 值域是“Class: 处理方法”。那么依据规则应该被转换成下面的形式:

$\langle ? x, \text{Property: troubleshooting}, ? y \rangle$

$\langle ? x, \text{rdf:type}, \text{Class: 故障描述} \rangle$

$\langle ? y, \text{rdf:type}, \text{Class: 处理方法} \rangle$

规则 5 若概念属于 ICSet 集, 约定查找的是该实例的所有属性及对应的值。

例如有实例“Individual: 滑环发热故障”。那么依据规则应该被转换成下面的形式:

$\langle \text{Individual: 滑环发热故障}, ? x, ? y \rangle$

规则 6 若概念属于 PICSet 集, 将该实例与属性进行合并转换。

例如: “Individual: 滑环发热故障”是“Class: 滑环故障”的实例, 而属性“Property: troubleshooting”在本体中定义域是“Class: 故障描述”, 值域是“Class: 处理方法”, 并且类“Class: 滑环故障”是“Class: 故障描述”的子类。那么依据规则应该被转换成下面的形式:

$\langle \text{Individual: 滑环发热故障}, \text{Property: troubleshooting}, ? x \rangle$

$\langle ? x, \text{rdf:type}, \text{Class: 处理方法} \rangle$

需要注意的是: 如果实例匹配到的是定义域类, 那么变量的位置应该在值域上; 反之, 如果实例匹配到的是值域类, 那么变量的位置应该在定义域上。

查询模式构造在查询语句构造器中占有很重要的地位。经过查询模式的构造, 这些三元组就可以被转换成与之对应的 SPARQL 查询语句。

(3) 构造查询语句

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^[7]。利用这些查询模式, 就可以完成对查询语句的构造^[8]。查询语句的构造示意图如图 3 所示。

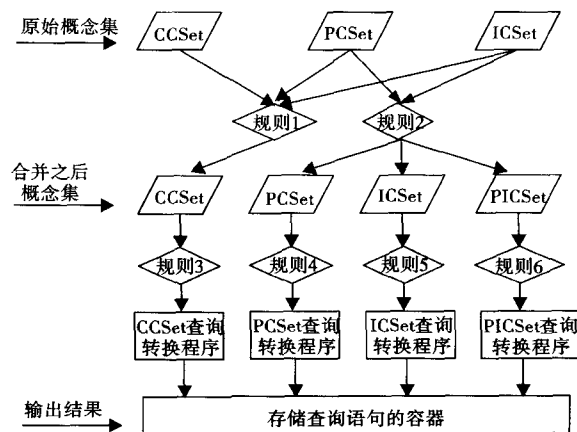


图 3 查询语句构造示意图

从图 3 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 3 至规则 6 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 检索引擎的实现与应用

针对所提出的矿井提升机故障知识语义检索的模型,本文设计并实现了一个基于本体和规则的矿井提升机故障知识检索系统.系统采用传统的基于 B/S 模式的三层体系结构,主要包括:显示层、中间层与数据层.在该层级结构中,显示层负责系统与用户的交互,结果的显示;中间层承担了接收显示层递交的事务,并做好事务的处理最终将处理结果返回到显示层等工作,是系统的核心模块;数据层负责数据的存储任务.系统层级结构如图 4 所示.

依据以上对查询构造器的分析,系统中查询构造器的类图如图 5 所示.其中包括三个查询词概念扩充实体类:ConceptConcept、PropertyConcept 和

IndividualConcept. 四个概念集合类:CCSet、PCSet、ICSet 和 PICSet. 还包括一个查询词类 QueryWord 和查询构造器类:QueryConstructor. 规则管理类图主要包括:规则实体类 Rule、规则库类 RuleBase 和规则库管理器 RuleBaseManage 类.系统通过这几个类的相互配合完成规则的编辑,规则库的建立及修改等操作.系统中与规则管理相关的类如图 6 所示.

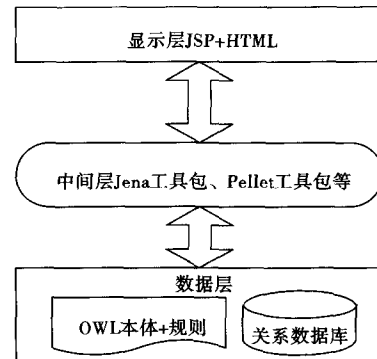


图 4 HFRS 系统层级结构

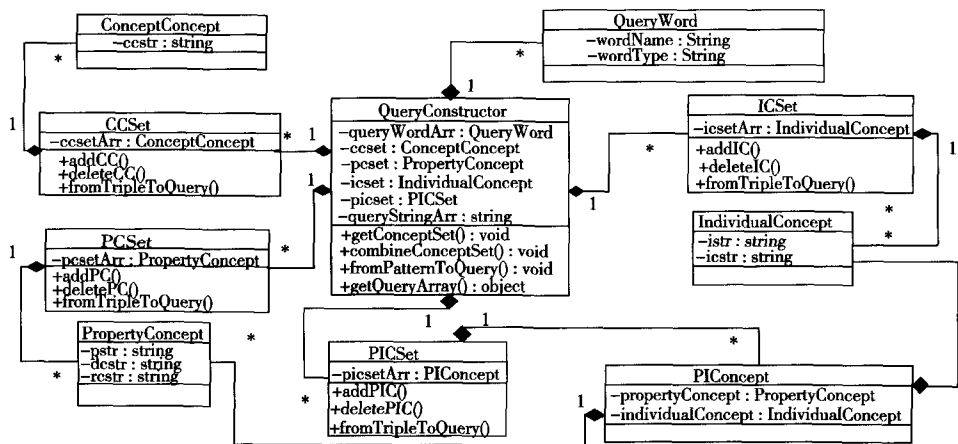


图 5 查询构造器类图

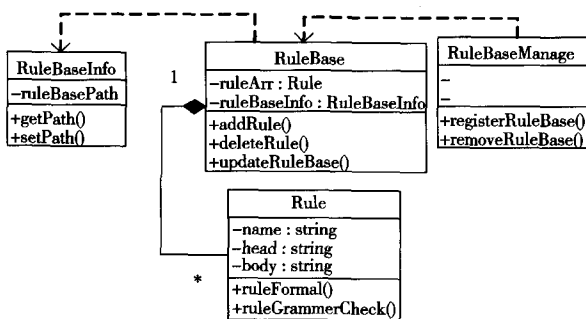


图 6 规则管理类图

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 Server2000
- 第三方包工具: JDK1.5.0 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.

5 结束语

本文从基于本体的知识检索系统结构出发,分析构建了一个基于本体知识库的提升机故障知识检索系统.通过原型系统的实现和应用,验证了本文所提出的方法的有效性和正确性,为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.

参考文献:

- [1] 马文峰,杜小勇.知识检索研究[J].情报理论与实践,2006(2):157-161.
- [2] 杨良斌,黄国彬,周静怡.近两年来国外有关本体基本问题的主要研究述评[J].图书馆建设,2008(8):80-83.
- [3] 丁晟春,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.
- [4] 许浩.基于本体论的语义检索技术的研究与实现[D].南京:南京理工大学,2007.
- [5] David Vallet, Miriam Fernandez, Pablo Castells. An

ontology-based information retrieval model[C]// Extend Semantic Web Conference, Greece, Aeraklion, 2005: 455-470.

- [6] 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.
- [7] Xiao Hang Wang, Tao Gu, Da Qing Zhang, et al. Ontology based context modeling and reasoning using OWL[C]// IEEE International Conference on Pervasive Computing and Communication. England: Nottingham, 2004.
- [8] 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:



许晓林 (1968—),高级工程师.研究方向为大型机电设备技术管理.

(上接第 63 页)

- [3] 宋相法,陈志国.基与 Fuzzy c-means 算法聚类有效性函数的纹理分割[J].河南大学学报:自然科学版,2004,34(1):14-17.
- [4] Xie X L, Beni G. A validity measure for fuzzy clustering [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1991, 13(8): 841-847.
- [5] Open_video project[EB/OL]. [2007-03-20] <http://www.open-video.org>.

作者简介:



冯燕女,(1986—),硕士研究生.研究方向为人工智能和图像处理.

王洪元 男,(1960—),博士,教授.研究方向为模式识别与智能系统.

程起才 男,(1981—),硕士,讲师.研究方向为机器学习和数字图像处理.