第37卷 第16期 计 算 机 工 程 2011年8月
Vol.37 No.16 Computer Engineering August 2011

·软件技术与数据库· 文章编号：1000—3428(2011)16—0054—03 文献标识码：A 中图分类号：TP301.6

基于本体的关联知识可视化检索模型

江潇俊，李善平，刘思屹

(浙江大学计算机科学与技术学院，杭州 310027)

摘 要：本体作为共享概念体系的形式化描述，在知识检索方面可解决海量知识利用问题。为此，在已有研究成果的基础上，提出一种基于本体的关联知识可视化检索模型。该模型从实用角度出发，关注知识源之间的关联性和知识检索的用户体验，改进传统的本体构建及维护方法，提出新的知识检索方法。应用实例结果表明，该模型能够提升用户获取知识的效率和质量。

关键词：知识检索；本体；本体构建；关联知识；可视化

Ontology-based Related Knowledge Visualization Retrieval Model

JIANG Xiao-jun, LI Shan-ping, LIU Si-yi

(School of Computer Science & Technology, Zhejiang University, Hangzhou 310027, China)

【Abstract】Ontology as a formalized description of shared conceptual system, can solve the problem of massive knowledge usage in knowledge retrieval. This paper suggests an ontology-based related knowledge visualized retrieval model based on full study of current research. From a practical point of view, the model focuses on the relationship between knowledge sources and user experience of knowledge retrieval, improves traditional ontology construction and maintenance method, and suggests a new knowledge retrieval method. The model can largely improve efficiency and quality of knowledge retrieval by users.

【Key words】knowledge retrieval; ontology; ontology construction; related knowledge; visualization

DOI: 10.3969/j.issn.1000-3428.2011.16.018

1 概述

随着知识经济时代的到来，各领域的知识资源库越来越大，与此同时新知识的创造速度也越来越快，如何有效地利用海量知识成为学术界和产业界共同关注的话题。例如：企业在探索如何更有效地利用知识资产，为其创造更多的利润；图书馆在探索如何更有效地组织图书知识，提升其使用效率和社会价值；教育行业在探索如何更有效地传播知识，提升教育的效率和质量。

很多知识利用的方法被提出并应用到了实践中，其中本体作为共享概念体系的形式化描述，被认为是解决海量知识利用问题的最重要的途径，知识检索是本体的一个重要应用方向。然而当前基于本体的知识检索模型存在 2 个主要的问题：(1)没有充分利用知识源之间、本体和知识源之间的关联性。(2)缺乏可视化和交互性，用户体验较差。这些问题使得用户无法了解到知识之间的关联，进而缺乏对领域知识全貌的认识，影响了知识利用、组织和传播的效果。

本文旨在解决这 2 个问题，从实用角度出发，提出一种基于本体的关联知识可视化检索模型。该解决方案的中心思想为：结合知识源的显性知识和领域专家头脑中的隐形知识构建和维护领域本体；当普通用户检索关键词时，向其提供关联知识图和知识源检索结果；用户可以通过关联知识图在相关知识及知识源之间快速导航。

2 相关工作

基于本体的知识检索模型在资源对象的组织、描述、表示、检索和模型约束等方面都具有自己的特征[1]。

在检索对象的组织上，知识检索模型利用领域本体作为组织资源的基础。首先构建一个涵盖相关领域概念及概念间关联的领域本体库作为资源描述和知识表示的工具与模型，如各学科领域的主题词表、分类表，在此基础上确定领域知识本体的主要概念和概念间的各种关系，构筑领域本体的概念模型。

在检索对象的描述上，知识检索模型借助语义标引工具，按照领域本体的概念及关联，对资源对象进行概念分析、分类、标引、描述和处理，形成机器可以理解的带有语义信息的元数据。

ODKM 系统提出了一种知识管理和检索的框架[2]。

基于本体的知识检索模型依赖于本体描述语言。本体描述语言用于对本体模型进行描述，通常使用 XML 语言。由 W3C 主持制定的 RDF 和 OWL 语言已经成为本体描述语言的事实标准。

基于本体的知识检索模型的重要步骤是领域本体的构建。本体构建过程不同于软件工程过程，没有统一的方法论。已经开发出的典型本体以及方法论包括 Cyc 本体及方法、企业本体及 Uschold&King 方法、TOVE 本体及 Gruninger&Fox 方法、Kactus 及 Bemaras 方法、Chemicals 本体与 Methontology 方法、Sensus 本体及方法。另外，还有很多致力于本体的自动化或半自动化构建的方法[3]。

本体的可视化旨在将本体模型用图形化的方式展示出来，以供领域专家和用户查看。可视化的方法主要包括：Indented list，Node-link and tree，Zoomable，Space-filling，Focus + context or distortion 和 3D Information landscapes[4]。

3 模型框架

本模型的框架采用分层设计的原则，有利于降低系统模

作者简介：江潇俊(1988－)，男，硕士研究生，主研方向：知识管理；李善平，教授、博士生导师；刘思屹，学士

收稿日期：2011-01-24 **E-mail：**jxj.jiang@gmail.com

块之间的耦合度、增加系统的扩展性和可维护性。模型框架包含 4 个层次，从下往上分别为数据资源层、基础服务层、业务逻辑层和知识表现层。图 1 为模型框架概览。

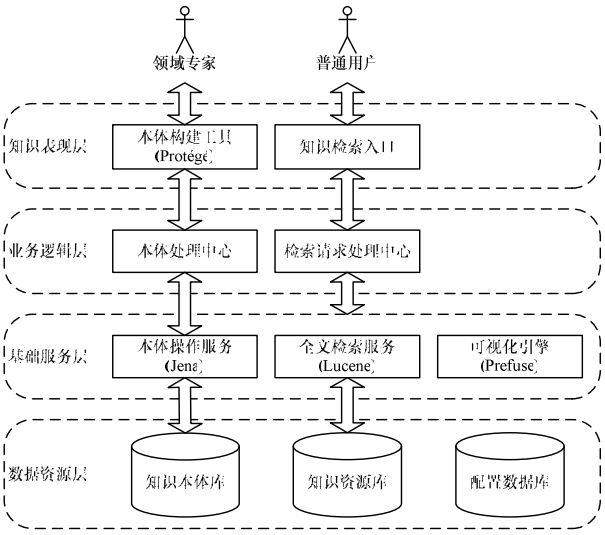

图 1 模型框架

3.1 数据资源层

数据资源层在文件系统或关系数据库中存放本模型所需的数据资源，主要包括知识本体库、知识资源库和相关配置数据。

知识本体库包含所考察领域的知识本体，可以以 OWL 格式存放在文件系统中，也可以存放在关系数据库中。前者易于查看、修改和维护；后者能够提高本体使用的性能，具体的策略依赖于性能方面的考虑和系统整体的设计。

知识资源库是指所考察领域的知识源，可以包含结构化数据、半结构化数据和非结构化数据。文本格式的知识源可以直接使用，而非文本格式的知识源则需通过相应工具进行格式转换和文本抽取，这方面的研究很多，如对于非结构化数据管理的研究。

配置数据库包本模型的配置参数，以 XML 格式存放。

3.2 基础服务层

基础服务层封装了底层标准化操作，为上层应用提供各种基础性服务，主要包括本体操作服务、全文检索服务和可视化引擎。

本体操作服务提供知识本体库的存取服务，将领域本体的存储细节与上层应用隔离。该服务使用 Jena API 操作领域本体。Jena 是由 HP Labs 开发的 Java 工具包，用于语义网应用程序开发，在业界很有影响力。Jena 一方面能够便捷地将本体存放在数据库中，便于本体的构建和维护；另一方面提供了强大的 API 对本体做各种操作，符合本模型本体操作的要求。

全文检索服务用于知识资源库的全文索引。本模型使用 Lucene 作为全文检索服务提供方。Lucene 是一个全文索引引擎的 Java 工具包，可以方便地嵌入到各种应用中实现针对性的全文检索功能。Lucene 作为全文检索领域最著名的开源框架，能够很好地为本模型提供服务。

可视化引擎提供丰富的基础可视化元素为数据提供直观的展示，是本模型实现良好用户体验的关键所在。本模型使用 Prefuse 作为可视化引擎。Prefuse 是一个 Java 编写的用户界面包，用来把有结构或无结构数据以具有交互性的可视化图形展示出来。Prefuse 提供的数据视图不仅非常直观，而且拥有丰富的视觉效果和很强的交互性，符合本模型的要求。

3.3 业务逻辑层

业务逻辑层是本模型的核心，接受前端知识表现层发送的各种处理请求，利用基础服务层提供的各种服务响应请求，并将结果返回给前端。主要的业务逻辑包括本体处理中心和检索请求处理中心。

本体处理中心负责本体的构建和维护操作，利用基础服务层提供的本体操作服务响应各种操作请求。

检索请求处理中心负责知识检索请求的处理，利用基础服务层提供的本体操作服务获取检索关键词的关联知识，利用可视化引擎生成关联知识图返回给前端。与此同时，利用全文检索服务对知识资源库进行检索。

3.4 知识表现层

知识表现层是本模型的前端，用于用户和模型之间的交互，主要包括本体构建工具和知识检索入口。

本体构建工具提供给领域专家构建和维护本体，本模型使用 Protégé 作为本体构建工具。Protégé 是由斯坦福开发的本体构建和知识获取系统。Protégé 拥有非常人性化的界面和强大的扩展机制，最大程度地减少领域专家构建和维护本体的困难。

知识检索入口提供给普通用户进行知识检索操作。用户在搜索框输入关键词或者通过上一次搜索产生的关联知识图点击关键词，该搜索词将被发送到业务逻辑层进行处理，并将处理产生的关联知识图和匹配的知识资源返回给用户。

4 模型关键流程

本模型包括 2 个关键流程，分别是本体构建及维护和关联知识可视化检索。

4.1 本体构建及维护

传统的本体构建及维护方法在相关工作中已有介绍，本模型在传统方法的基础上进行改进，提出了下述的本体构建及维护方法：

(1)确定知识的范围。该步骤的首要目标是确定考察领域的明确范围，否则知识的形式化表述就无从谈起。在该步骤中，首先让领域工程师确定该领域内的知识范畴，包括一些概念及其之间的联系，然后提供一个关于这些知识概念的详尽描述并且根据知识量建立一个本体论的分类。

(2)知识本体的构建。该步骤的目的是获取上一步定义好的目标知识。在该步骤中，首先需要定义好用于获取目标知识的输入变量和来源，包括哪些隐形知识和显性知识。隐形知识的提取往往是通过和领域专家的头脑风暴进行的，显性知识的提取则可以借助半自动化的工具。

(3)知识本体的表达。知识在提取出来之后，就需要运用某种方式表达出来。一般采用描述逻辑(或称概念逻辑)的方法表达知识，基于一种声明性的形式化语言。常用的标准是 OWL 语言，在此不再详述。

(4)知识本体的处理和存储。知识本体在构建成功之后，需要利用诸如 Jena API 将其存入到相应的数据库中，以便之后的应用。

4.2 关联知识可视化检索

大多基于本体的知识检索模型并没有使用关联知识的概念[5]，而这正是关联知识可视化检索的切入点。

关联知识可视化检索的核心是关联知识图，该图是以一个本体概念或实例为中心，多个关联概念和实例为端点的罗盘。其具有的特征包括：用连接线展示概念和实例之间的关

联性；用不同的可视方式(如颜色)区分概念和实例以及各种不同的关联性；概念和实例可用鼠标点击，点击后将该概念或实例作为检索关键词重新检索。此外，关联知识图需拥有丰富的视觉效果和良好的用户体验。

图 2 为检索请求处理中心的流程。

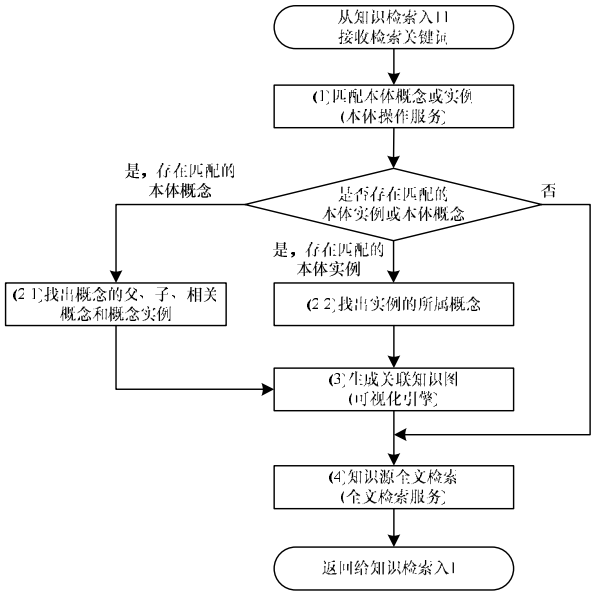

图 2 关联知识可视化检索流程

检索流程各步骤说明如下：

(1)匹配本体概念或实例。首先将本体模型读入到内存中，然后寻找与检索关键词匹配的本体概念或实例，这些操作都需要利用基础服务层的本体操作服务。若存在匹配的本体概念，则执行步骤(2.1)；若存在匹配的本体实例，则执行步骤(2.2)；若不存在匹配的本体概念或实例，则直接跳到步骤(4)。

(2.1)找出概念的父、子、相关概念和概念实例。利用基础服务层的本体操作服务分别找到匹配概念的父概念、子概念、相关概念和概念实例，以方便下一步生成关联知识图。

(2.2)找出实例的所属概念。利用基础服务层本体操作服务找到匹配实例的所属概念，以方便下一步生成关联知识图。

(3)生成关联知识图。利用上一步找到的概念和实例的关系生成如前所述的关联知识图，这些操作是利用基础服务层的可视化引擎实现的。

(4)知识源全文检索。类似于普通的全文检索，利用基础服务层的全文检索服务进行，将检索结果和关联知识图共同返回给前端。

这种检索方法通过关联知识图体现了知识源之间、本体和知识源之间的关联性，并提升了用户体验。

5 操作系统网络教育平台应用实例

本模型解决了原有知识检索模型中存在的问题，在诸如企业知识库、数字图书馆和网络教育等领域有着广阔的应用场景。本文以操作系统网络教育平台为例说明如何应用本文提出的模型，并展示应用的效果。

操作系统网络教育平台是浙江大学精品课程《操作系统》的日常教学辅助平台，不仅用于教师和学生之间的交流沟通，并且真正融入到教学和考核当中，极大地方便了学习和教学工作。

5.1 操作系统本体构建

简单来说，操作系统本体的层级可以分为 2 层，如图 3 所示。上层为操作系统本身，底层则是操作系统内部运行的各种概念，如内核、进程管理、存储器管理等。上层主要包含的是现在存在的不同类型的操作系统，如实时操作系统、批处理操作系统等，它将操作系统这个整体进行了分类，将现存的使用不同技术和原理的操作系统进行了整理。底层反映了操作系统中的内部结构，通过底层的概念，能够分解操作系统，让各种关于操作系统修改、运行和操作的信息通过这一层得到很好的整理和分类，同时又能在一些表面看似不相关而实际内容有关的概念之间构建联系。另外，用户在层级中处于一个特殊的位置，他不属于任何层级，但他是负责操作和输入指令的。

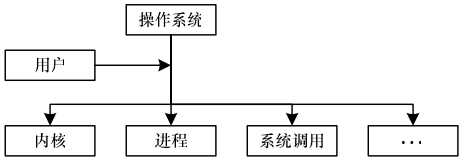

图 3 操作系统本体层次

本文利用模型中提出的本体构建和维护方法创建操作系统本体模型，共包含 53 个本体概念、322 个本体实例和 23 个本体属性。部分本体模型如图 4、图 5 所示。

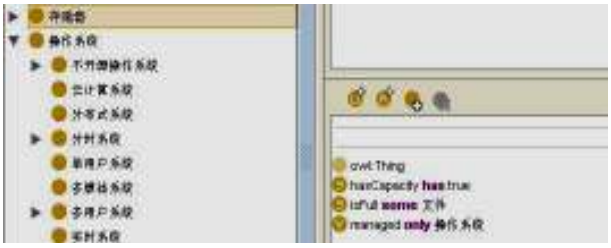

图 4 部分本体模型 1

图 5 部分本体模型 2

5.2 关联知识可视化检索实现

本文利用模型中提出的关联知识可视化检索方法实现知识检索入口。图 6 所示为查询嵌入式系统时的截屏。

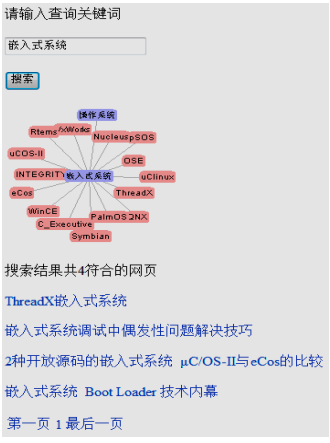

图 6 查询嵌入式系统时的截屏

(下转第 59 页)