

本体存储技术研究

乐 超, 王晓军

(南京邮电大学计算机学院, 江苏省南京市 210003)

摘 要 当前, 基于本体的网络管理是研究网络自动配置和管理的研究热点。在基于本体的网络管理中, 如何实现本体的高效存储是十分关键的问题。文章分析本体存储技术的研究现状, 比较本体的几种主要存储模式, 指出它们各自的优缺点和应用范围。在讨论当前本体存储模式的基础上对该技术的发展方向进行了展望。

关键词 本体论; 语义; 网络管理; 数据存储

当前, 网络管理技术正向自动化和智能化的方向发展。本体因其富含语义信息, 能够被机器理解而被认为是实现网络管理自动化和智能化的有效途径。在基于本体的网络管理技术中, 实现本体的高效存储和查询是个十分关键的问题。本文分析比较了本体的几种主要存储方法, 指出了它们各自的优缺点和应用范围。最后, 在分析总结当前本体存储模式的基础上对该问题的发展方向进行了展望。

1 本体论概述

本体是对特定领域重要概念明确的、形式化的描述。明确的是指被描述的概念值应有所限制(如年龄), 此概念的值应为一数值(int), 而非字符串(string)。形式化指的是概念的描述具有标准格式, 能被计算机理解。本体主要构成要素包括类(class)、slot、实例(instance)和公理(axiom)。类是人脑中的一个类别或概念, 如防火墙就可作为一个类。slot 用来描述概念的属性或概念之间的关联, 概念间的继承关系也属于 slot, 如防火墙有硬件防火墙和软件防火墙。实例是类的一个具体案例, 它继承了类的所有属性和关联。公理定义了概念间的管理或限制。与 slot 不同之处在于 slot 明确定义了两个概念之间的关联, 公理则定义了那些 slot 无法定义的关联, 这种关联往往会跨越多个概念。

本体是面向计算机的, 因此在描述手段上强调明确无歧义。近年来, 本体描述语言主要出现了 RDF、RDFS、OL、DAML-OL、OL+DAML 和 OWL 等。这些本体描述语言的共同特点是它们都是基于

XML 语法的, 实质上就是 XML 文档。

2 本体存储的主要方法

目前, 本体的存储方法主要有文本存储和数据库存储两种。

2.1 文本存储方式

文本存储方式是将本体库以文件形式存储在本地文件系统中。采用该方式, 系统需要将本体库从文件读入内存中, 然后在内存中对本体库进行操作, 最后将本体库再写回文件。本体库可采用多种文件格式: 如 XML、RDF、OWL 等。

文本存储具有以下优点: a) 文本文件能完整保存本体的语义信息, 具有较好的可扩展性和灵活性。b) 已有若干不断发展中的针对 OWL 等文件的查询和推理工具, 如 jena、protégé 等本体推理机。文献[3]对 OWL 的本体存储模式进行了分析。c) 本体定制和管理简单方便, 许多本体编辑管理工具能直接生成 OWL 文件(如 protégé 本体编辑工具), 但是这种方法也有一个明显的缺点, 效率不高, 难以适应数据量较大的情况。在该系统中, 对本体的每一次操作都会涉及到对整个本体文件的操作。尤其是当本体规模较大时, 系统需要进行大量的内存管理, 效率难以得到保证。在实际应用中, 基于文件系统的本体存储往往只适用于本体的编辑和建立, 而不适用于大规模的本体数据存储和查询。

2.2 数据库存储方式

数据库存储方式是将本体按照一定的策略组织存放在数据库中。该方法利用了数据库在管理和存

储数据方面的优点,提高了本体存储和查询的效率。由于关系数据库技术发展成熟,当前许多关于本体存储方面的研究都将关系数据库作为本体的存储和管理系统。但由于大多本体描述语言(如 RDF、OWL 等)都是基于 XML 的,利用 Native-XML 数据库存储本体信息等相关研究也得到了越来越多的关注。文献[4]详细介绍了 XML 数据的存储模式。

2.2.1 本体到关系数据库的存储模式

关系数据库并不是适合表达本体的好方式。本体含有语义信息,其逻辑性比关系模式复杂得多。利用关系数据库作为本体的存储后台需要对本体的存储模式进行仔细设计,在尽量不丢失本体语义信息的情况下提高本体存储和管理的效率。经过多年的研究,本体到关系数据库的存储模式主要可分为水平表模式、垂直表模式和分解模式。

1) 水平表模式

该模式下数据库中只有一张通用表,表中的列是本体的属性,本体有多少属性表就有多少列。本体的每个实例都是表中的一条记录。该模式中,表包含了大量字段,查询效率高。但也存在以下几个问题:a)表的字段过多。若是本体较大,则表的字段可能会超出关系数据库所允许的范围。b)空间冗余较大。本体的一个实例可能只有少数几个属性,却也要占据表的一行,把没有用的属性设置为空,这样带来了大量的空字段,造成较大的空间冗余。c)可扩展性较差。一旦本体发生变化,则会造成表模式的改变,这样的改变代价很大。

2) 垂直模式。

在该模式下,数据库以三元组表来存储本体。表中的每条记录对应于一个 RDF 三元组。这种模式设计简单,结构稳定,修改本体时只需修改表中相应的元组。但是与水平模式相比,垂直模式查询效率较低,对每个本体的查询都需要搜索整个数据库,尤其是那些需要表连接的查询效率非常低。另外,该模式下的查询语句较复杂,容易出错。

3) 分解模式

该模式的基本思想是对数据库进行模式分解,根据分解对象的不同,分为两种方法:a)基于类的分解方法。该方法为每个类都创建一张表,表名为类名,表的列为类的属性。这种表结构清晰,但若本体中类或属性发生变化时表结构也要改变。b)基于属性的分解方法。该方法为每个属性创建一张表,表名

为属性名,每个表包含两列,对应 RDF 中三元组中的 subject 和 object。该分解方法中,对类的隐含实例查询代价较大。这两种分解模式方法随着本体的变化都要不断的创建和删除表,代价较大。

以上几种本体存储模式都各有优缺点,仅采用一种模式并不是好的选择。当前基于关系数据库的本体存储系统中,大多混合采用了以上几种模式并结合应用环境加以改进,以尽可能提高系统的效率。文献[2]针对 OWL 本体类和属性的特点,通过单独设立类关系表、添加关系约束表等方法改进已有模式,提高了本体查询的效率。

2.2.2 基于 Native-XML 数据库的存储方式

由于 OWL 等绝大多数本体描述语言都是基于 XML 的,OWL 本体通常都是以 XML 格式描述和保存的。许多本体工具都支持对 XML 文件格式的本体进行操作。

Native-XML 数据库以 XML 文档作为其基本存储单位,数据库内部的数据模型是依据 XML 特性建立的,而不是关系表^[1]。利用 Native-XML 数据库对 XML 文档进行保存、查询和更新可以提高系统效率,也能降低实现的复杂度。

与基于文件系统的存储方式相比,将 OWL 本体以 XML 文档的格式存储在 Native-XML 数据库中,不但具有前者所具有的那些优点,还弥补了前者在查询效率上的不足。与关系数据库相比,以 Native-XML 格式存储 OWL 本体具有以下几个优势:a)对本体的描述能力较强,能完整保存本体所具有的语义信息。b)能直接以 XML 格式对本体进行操作,操作简单方便,无需进行数据结构等的转换。c)无需对本体进行分解,系统设计简单。d)支持对具有动态结构的数据和半结构化数据的处理,系统的可扩展性和灵活性较强。文献[5]介绍了如何利用 XML 数据库 xindice 存储本体库,设计了一种本体查询方法。

2.3 几种本体存储方式比较

表 1 对上述几种本体存储方法的优缺点及其应用情况做了简单总结。

3 本体管理工具介绍

3.1 Protégé

Stanford 开发的 Protégé 是目前使用最广泛的本体编辑工具之一。Protégé 树形目录结构显示本体

表 1 三种存储方式比较

存储方式	优点	缺点	应用现状
文本存储	系统简单快捷,管理方便,可扩展性好。能完整保留语义信息。具有较好的工具支持。	系统效率较低。	主要用于本体的编辑和建立及小规模的本体存储管理系统。
关系数据库	效率高,技术成熟。能充分利用 SQL 语言的优势。	系统设计复杂,可扩展性和语义支持较差。	目前大规模本体数据存储管理的主要方式。
Native XML 数据库	系统设计简单,可扩展性较好。能完整保存语义信息。系统效率高。潜力巨大。	技术没有关系数据库成熟。	将来的发展方向。

中的类和属性,允许用户增加或编辑类、子类、属性、实例等,允许用户在概念层次上进行领域模型设计。Protégé 具有很强的可扩展性,通过安装不同的插件取得对本体的不同操作能力。除了提供本体编辑工具以外,Protégé 还提供了一整套的 OWL API。该 API 提供了加载和保存 OWL 文件、查询和操作 OWL 数据模型以及执行基于描述逻辑的推理的类和方法。Protégé 支持多种文件格式导出本体,如 RDF、RDFS、OWL、XML Scheme 等。它还可以利用 JDBC 将本体存储到关系数据库中。

3.2 Jena

Jena 是来自于 HP 实验室的开源系统,它提供了本体的创建、查询、推理、存储等功能。Jena 的 RDF API 允许处理基于 RDF 的本体数据,支持 OWL、DAML+OIL 和 RDFS 等本体描述语言。Jena 的本体子系统支持对导入本体的文档管理,推理子系统能从本体数据中获得隐含的信息。Jena 提供了 ARQ 查询引擎,支持 SPARQL 本体查询语言或 RDQL 查询语言对本体模型进行查询。Jena 支持以文本文件或关系数据库的方式存储本体,它提供了将基于 RDF 的本体数据存入关系数据库的接口。在实际操作中,应用程序不必直接访问数据库,也不必知道数据库的模式,只需通过 Jena 提供的接口访问本体数据。目前 Jena 实现了对 MySQL、HSQldb、PostgreSQL、Oracle 和 SQL server 等关系数据库的支持。

3.3 Sesame

Sesame 也是一个开源的项目,它是一个能为 RDF 和 RDF Schema 中的大量数据提供有效存储

和丰富查询能力的框架。Sesame 提供了非常开放的 API 接口,可以很方便地集成不同的存储系统、推理引擎和查询引擎等功能模块。Sesame 的设计和实现独立于任何存储设备。因此,Sesame 能被部署在各种各样的存储设备上(如关系数据库、对象数据库、XML 数据库等),而不需要更改查询引擎或其他的功能模块。Sesame 的系统框架如图 1 所示。

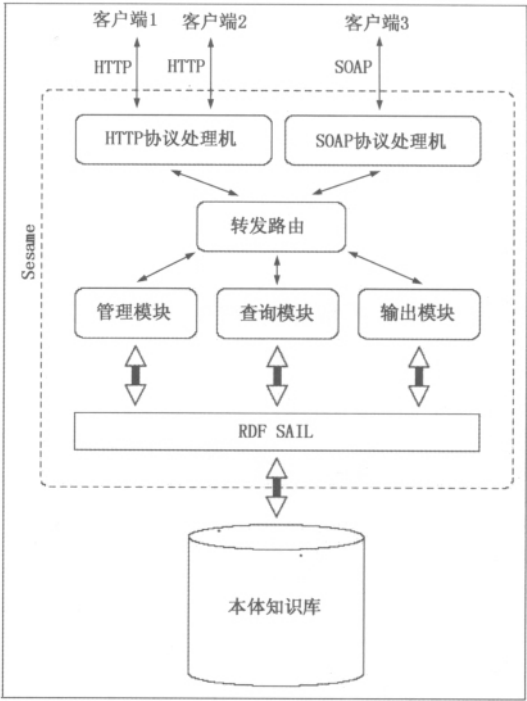


图 1 Sesame 系统框架

4 总结与展望

现有的本体存储方法还不能很好解决本体存储

管理问题。基于文本的存储方法效率太低,只适合小规模本体存储解决方案。基于关系数据库的本体存储虽然提高了系统的效率,但仍然存在诸多问题。关系数据库并不是针对本体的特点设计的,关系数据库的表结构不能很好表示本体数据结构。基于关系数据库的本体存储需要将复杂的本体图分解成简单的关系存储,将基于图的查询转换成大量关系查询。系统设计复杂,可扩展性和适应性较差,大大限制了关系数据库处理大规模本体数据的能力。同时,由于关系数据库不能描述语义信息,基于关系数据库的本体存储往往预先将本体中隐含的数据转化为显示数据存储,这种做法虽然提高了查询效率,却容易造成隐含信息的丢失。当更新原有数据时,很难保证本体的隐含信息与转换后的显式信息之间的一致性。

从本文的分析来看,基于 XML 数据库是本体存储问题的一个很好的解决方案。在以后的研究中,

为基于 XML 数据库技术设计一种专门的本体存储系统是一个很值得重点研究的解决方案。

参考文献

- 1 陈光仪,徐德智. RDFS 本体在关系数据库中的存储研究[J]. 计算机数字工程,2008(12):181-184.
- 2 陶 皖,姚红燕. OWL 本体关系数据库存储模式设计[J]. 计算机技术与发展,2007(2):111-115.
- 3 李 洁. OWL 本体存储模式研究[J]. 中国科技信息,2007(21):118-121.
- 4 邓华梅,等. 关于 XML 数据的存储研究[J]. 科技情报开发与经济,2008(24):153-155.
- 5 沈洪良,朱国进. 基于 Xindice 的本体存储查询研究[J]. 计算机应用研究,2005(12):36-39.

乐 超(1983—),男,硕士研究生,主要研究方向为分布式网络。

收稿日期:2009-05-18

(上接第 19 页)

网络优化前周忙时平均反向软切换成功率为 80.36%,而网优后的指标为 98.91%,网络优化后的指标比网优前提高了 18.55 个百分点,网络优化实施效果明显。

4 结束语

网络指标监控是在日常网络优化过程中的重要内容,在日常网优中应加强 TOPn 分析,重视基础优化工作,把网络优化工作与网络维护有机结合在一起。EVDO 软切换成功率优化工作应重点关注 EVDO 反向软切换成功率在 98%以下的 BSC (基站控制器);再具体定位到问题扇区。在日常优化中应关注新开通的扇区,在进行闭锁扇区时,基带增益最好设置为 0,或直接关闭功放电源,以避免闭锁扇区仍

然有小功率输出而导致软切换失败的事件发生,引起 EVDO 软切换成功率低下,影响用户感知度。在 EVDO 邻区优化工作中,配置邻区时尽量避免配置闭锁邻区。另外,还需要控制扇区覆盖范围,尽量避免越区覆盖导致同 PN 干扰。规划 PN 时应充分保证 PN 的复用距离。

EVDO 网络建设时间短,EVDO 网络优化在目前基本上处于探索阶段,很多优化的经验需要在日常网络优化的过程中不断积累和完善,目前对于 EVDO 反向软切换的研究更是少之又少,本文旨在抛砖引玉,希望对 EVDO 网络优化工作有所帮助。

吴奕生(1964—),男,高级工程师,主要从事 CDMA 无线网络优化等方面的工作。

收稿日期:2010-04-12