

P2P 中基于本体论的知识管理框架模型及实现

孙宝军 王新军

(山东大学计算机科学与技术学院 济南 250061)

摘要 为了将 P2P 中对等体的资源描述为结构化的知识,以提高资源共享,本文提出一个基于本体论的知识管理框架模型。从知识的建立和知识的检索两方面分析了模型的实现技术。探讨了相互协作的对等体按预定义查询模式和本体论匹配的语义技术实现信息搜索和知识获取的过程。

关键词 本体论,语义 Web,知识管理,P2P

A Framework for Ontology-Based Knowledge Management in P2P System

SUN Bao-Jun WANG Xin-Jun

(School of Computer Science and Technology, Shandong University, Jinan 250061)

Abstract In order to improve resources sharing, resources of Peer be described to structural knowledge. In this paper a general framework is proposed for ontology-based knowledge management. After that, the implementation technique of this model is analyzed from two aspects, knowledge establishing and knowledge searching. Peer ontologies describe the knowledge of each peer, and on collaborations among peers, allowing information search and knowledge acquisition, according to pre-defined query models and semantic techniques for ontology matching.

Keywords Ontology, Semantic Web, Knowledge management, Peer-to-Peer

1 引言

Peer-to-Peer(P2P)网络是一种分布式系统,系统中每一个直接与其它节点交换信息和服务的节点(也叫对等体)具有相同的角色和能力。许多 P2P 系统,如 Napster、Freenet 和 Gnutella 等的应用较好地解决了以对等体为中心的相互协作和资源共享问题。但传统的对等体之间通常利用简单的文件名作为元数据进行文件交换以实现资源共享,由于对等体中的资源大多是非结构化的,仅用文件名不足以描述资源特征。为了有效实现资源共享,在此采用语义 Web 技术实现对对等体数据和资源的语义描述。语义 Web^[5,6]的目标是使得 Web 上的信息具有计算机可以理解的语义,满足智能软件代理(Agent)对网络上异构和分布信息的有效访问和搜索。其核心层为 XML、RDF 和本体论。在语义 Web 中,本体论^[4]是解决语义层次上数据资源共享和交换的基础。

本文采用本体论对表达资源特征的元数据进行语义丰富的描述,并在此基础上提出一个基于本体论的知识管理系统框架。知识管理是指综合利用信息技术实现知识的建立、存贮和重用。重用的关键技术是如何有效进行知识的检索,即本体论匹配技术。具体的匹配算法采用 ARTEMIS 数据综合系统中提出的模式匹配技术^[1]。它是基于语言学的语义关系提出了模式匹配算法。与该系统不同的是本文应用了本体论来描述资源元数据语义关系。为此本文从知识的建立和知识的检索两方面分析模型实现的技术。探讨了相互协作的对等体按预定义查询模式和本体论匹配的语义技术实现信息搜索和知识获取的过程。

本文第 2 节提出一个基于本体论的知识管理系统框架,第 3 节提出查询处理技术,第 4 节描述本体论匹配技术。最后展望了此框架的应用前景。

2 基于本体论的知识管理系统框架

在知识管理系统中,要实现有效的知识共享必须对对等体的资源进行语义清晰的、概念化的形式描述以建立对等体的知识。本文采用对等体本体论描述每个对等体的知识,并在此基础上给出了知识管理系统框架。

2.1 对等体本体论结构

本体论的结构在不同的文献有不同的描述。本方法中,对等体本体论的结构按照概念、属性和语义关系组织成本体论知识,形式定义如下:

定义 1(PO , 对等体本体论) 一个对等体本体论 PO 定义成一个 3 元组的形式:

$$PO = (C, P, SR).$$

其中, C 是 PO 概念(concepts)的集合。其包含内容知识层的内容概念和网络知识层的网络概念。 P 是概念属性(properties)的集合。 SR 是内容概念之间语义关系(semantic relations)的集合。其包含同义关系,上/下位关系等。

2.2 知识管理系统框架

依据对等体本体论结构,本文提出如下的知识管理系统框架(见图 1)。

如图 1 所示,基于对等体本体论的知识管理系统框架结构分为 3 部分:

(1)对等体的数据:存贮在对等体的各种资源,即各种数据/信息(如文档,多媒体文件等)。

(2)对等体的知识:我们把存贮在对等体的资源按照对等体本体论结构组织成对等体的知识。也就是使用对等体,对资源的元数据按语义技术进行描述,将资源的非结构化信息转化为结构化的知识,存贮在对等体本体论的知识库中。这里也把对等体本体论概念化为 P2P 系统中的一个概念,这个概念

供了基本的通信原语,但它们没有实现更高层次的抽象。如果直接以它们为基础进行开发,结果往往是工作量大,程序容易出错,程序模块的封装程度低,可复用性差。

JAVA RMI、基于 WINDOWS 的 DCOM 以及跨平台的 CORBA 是目前应用比较广泛的中间件平台。开发人员可以按其指定的规则把应用程序的功能封装在一系列可复用的模块中,然后通过它们提供的基础平台实现模块间的通信交互。这大大提高了开发效率和系统的模块化程度。但它们往往对底层平台有较高的要求,也缺少 QoS、异步通信等特性。同时,作为一个通用平台,它们不可能针对某个应用程序作专门的优化。

STREAMS 是 UNIX 内核实现的一个框架。它被广泛应用于网络协议栈、设备驱动程序和终端服务的开发。CACF 借鉴了 STREAMS 的流的概念,运行时模块的压入与弹出和避免内存拷贝的复合消息结构。CACF 与 STREAMS 的不同之处在于,STREAM 中的所有模块都必须使用唯一的由系统提供的队列模型及基于其上的流控算法,而 CACF 允许应用程序实现自己的队列模型和流控算法。

CLICK 是为开发软件路由器而设计的一个框架。一个 CLICK 程序就是一个有向图。图的每个节点是一个被称为 Element 的程序模块,它实现某个具体功能。图的边代表 IP 报文在结点之间可能的流通过路。CACF 借鉴了 CLICK 把各种队列模型也作为独立模块实现,然后在完成实际报文处理的模块之间插入各种队列模块,从而实现 QoS、流控、带宽分配的思想。但 CLICK 缺乏消息复合机制,并且只支持协作过程这单一的并行模型。

(上接第 32 页)

在实际应用中采用全局相似度来全面评估两个概念相匹配的程度,它是上面相似度计算值的线性组合。对每种相似度引入一个适当的权值来计算全局相似度。在应用中可采取灵活的匹配策略。利用全局相似度的值,可以在一定程度上控制匹配目标概念的概念数量。给定目标概念 C 和对等体本体论 PO ,用 M_c 表示匹配 C 的 PO 的概念集合,即相对 C 有一个全局相似度值 $GA \neq 0$ 的概念。这里定义两种匹配策略用于选择匹配概念:

(1)相似:在相似策略里,其目的是发现与概念 C 语义对应的概念。所有全局相似度值大于等于相似域值 $t_s > 0$ 的概念被检索,即

$$M_{sc} = \{C' \in PO \mid GA(C, C') \geq t_s\}.$$

(2)等价:在等价策略里,其目的是发现等价的概念。所有全局相似度值大于等于等价域值 t_e ,且 $t_e > t_s$,这样的概念被检索,即

$$M_{ec} = \{C' \in PO \mid GA(C, C') \geq t_e\}, \text{ 其中 } M_{ec} \subseteq M_{sc}$$

等价策略是一个更严格的策略,其返回的值是相似策略检索概念的子集,这些概念与 C 有极高的相似度。

继续上面的查询例子,考虑图 2 的查询 B ,其指定了概念名字和概念属性,将其转换为 $BOOK$ 的本体论的描述。然后与对等体的本体论 PO_B 的每一个概念进行比较,估计名字和结构的相似度的系数。通常在 $Book$ 和 $Volume$ 之间会有一个较高的相似度值,因为它们的名字是同义关系,而且它们有两个公共的属性。而在 $Book$ 和 $Library$ 之间,由于没有公共的属性,其相似度就低。这里设定等价域值 $t_e = 0.9$ 和相似域值 $t_s = 0.5$ 用于全局相似度概念的选择。按照相似匹配策略,则

SEDA 是一个为开发高度并行的网络服务器而设计的框架。在 SEDA 中,程序的功能被划分到多个级连的 Stage。每个 Stage 由一个队列和多个从该队列中读取请求的线程构成。SEDA 能支持很高的并发度,同时保证服务器在超载情况下的稳定性。但由于线程调度等原因,它不适用于对服务实时性要求比较高的应用。

结论 本论文提出了一个支持 P2P 客户端程序开发的应用程序框架 CACF,并讨论了使用它开发的具有复杂通信需求的应用程序 JCVIEWER。程序的开发实践证明,CACF 灵活的扩展机制,水平、垂直和协作过程等多种并行模型,复合的消息结构,通过消息优先级和多种队列模型实现的流控和 QoS,以及多路复用等机制能有效地降低开发通信程序的复杂度,构造出健壮高性能的 P2P 客户端程序。

参考文献

- 1 Schmidt D. Applying Patterns and Frameworks to Develop Object-Oriented Communication Software. Handbook of Programming Languages, Vol. 1 MacMillan Computer Publishing, 1997
- 2 Schmidt D, Suda T. Transport System Architecture Services for High-Performance Communication Systems. IEEE Journal on Select Areas in Communication, 1993
- 3 Java 2 Platform, Standard Edition, v 1. 4. 1 API Specification, Sun Microsystems Press
- 4 STREAMS Programming Guide. Sun Microsystems Press
- 5 Welsh M, Culler D. SEDA: An Architecture for Well-Conditioned, Scalable Internet Services. The 18th Symposium on Operating Systems Principles, 2001
- 6 Morris R, Kohler E. The Click Modular Router. The 17th ACM Symposium on Operating Systems Principles, 1999

返回 *Volume* 和 *Journal*。按照等价匹配策略,则只返回 *Volume*。

结束语 由于本体论具有良好的概念层次结构和对逻辑推理的支持,在信息检索,特别是基于知识的检索中得到了广泛的应用。本文利用本体论较好地解决了 P2P 系统中对等体相互协作时的知识共享和积累。基于本文提出的本体论知识管理的框架结构,不仅能在纯 P2P 系统相互协作的对等体间实现知识共享和积累,而且在混合的 P2P 网络结构中,这种基于协作的方法也能用来实现超级节点之间的知识共享。

当然,该框架的有效实施还取决于对等体本体论知识的建立,这涉及到两个方面,一是对等体本体论的建立(这是一个由领域专家不断修改、完善的过程),另一个是如何用本体论对对等体的资源进行有效的描述。

参考文献

- 1 Castano S, Antonellis V D, Vimercati S D C D. Global viewing of heterogeneous data sources [J]. IEEE Transactions on Knowledge and Data Engineering, 2001, 13(2)
- 2 Castano S, Ferrara A, Montanelli S, Pagani E, Rossi G. Ontology-addressable contents in P2P networks [J]. In: Proc. of WWW'03 1st SemPGRID Workshop, Budapest, Hungary, May 2003
- 3 刘群,李素建.基于《知网》的词汇语义相似度计算 [J]. Computational Linguistics and Chinese Language Processing, Aug. 2002, 17(2): 59~76
- 4 Perez A G, Benjamins V R. Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem-Solving Methods [J]. In: Stockholm V R, Benjamins B, Chandrasekaran A, eds. Proc. of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods (KRR5) 1999. 1~15
- 5 Heflin J, Hendler J. A portrait of the Semantic Web in action [J]. IEEE Intelligent System, 2001, 16: 54~59
- 6 Berners-Lee T, Hendler J, Lassila O. The Semantic Web [J]. Scientific American, May 2001