

基于 Super-P2P 语义网络的知识共享模型

李志国^{1,2}, 冯永¹, 钟将¹, 叶春晓¹

(1. 重庆大学计算机学院, 重庆 400030; 2. 上海宝信软件西南研发中心, 重庆 400041)

摘要:在语义网络的研究中,如何超越集中式的语义存储,实现更高层次的资源共享,是目前语义网络的研究热点之一。在分析多种 P2P 结构的基础上,该文 Super-P2P 结构应用于语义网络,以解决语义网络的语义集中式存储问题。研究了 Super-P2P 语义网络的资源发现机制以及基于其的知识服务。应用案例分析证明,基于 Super-P2P 语义网络的知识共享模型可以实现更高层次的资源共享,提供高质量、高可靠性的知识服务。

关键词:知识服务; 超级 P2P; 语义网络; 小世界; 本体

Knowledge Sharing Model Based on Super-P2P Semantic Grid

LI Zhi-guo^{1,2}, FENG Yong¹, ZHONG Jiang¹, YE Chun-xiao¹

(1. College of Computer Science, Chongqing University, Chongqing 400030;

2. Southwest Research Center, Shanghai Baosight Software Corporation, Chongqing 400041)

【Abstract】Semantics distributed storage for better resource sharing is a focus in the study of semantic grid. This paper analyzes variety of P2P structures, and applies Super-P2P structure to semantic grid. Semantic grid based on Super-P2P is presented to address problems of semantics centralized storage in semantic grid, and resource discovery of Super-P2P semantic grid and knowledge services based on Super-P2P semantic grid are reported. Application case study demonstrates that the knowledge sharing model based on Super-P2P semantic grid can share resources better, and provide high-quality, high-reliability knowledge services.

【Key words】knowledge services; Super-P2P; semantic grid; small-world; ontology

语义网(semantic Web)是由文献[1]提出的下一代WWW技术,网格(grid)是由文献[2-3]提出的下一代分布式计算技术。两方面的交叉即语义网络(semantic grid)则是一个新兴的研究领域。“超越集中式的语义存储”是文献[4]提出的语义网络面临的巨大挑战之一。目前,对等网(Peer-to-Peer, P2P)^[5]技术因具有发挥闲散资源利用率特征而得到业界的关注。与集中式服务模式相比,P2P结构在可扩展性方面具有天生的优势,通过增加资源或者节点的组织和管理功能以实现资源共享。基于上述分析,将P2P技术和语义网络技术相结合研究,可以解决语义网络集中式语义存储的瓶颈,提供高质量、高可靠性的知识共享服务。

1 P2P 网络模型分析

从拓扑结构和技术上P2P网络可分为纯分散式P2P网络、混合P2P网络和超级P2P(Super-P2P)网络^[6]。

1.1 纯分散式 P2P 网络和混合 P2P 网络

纯分散式P2P网络中没有服务器,节点之间直接建立连接。它存在如下的问题^[5]:缺乏中心管理者,网络节点难以发现,安全性较差。典型代表是Gnutella, FreeNet等。

混合 P2P 网络的构建须通过服务器集中认证,并建立索引机制。相对于纯分散式 P2P 网络,它易于发现网络节点、易于管理且安全性较好,但由于服务器瓶颈的限制,规模不易扩大,容错性差。其典型代表是 Napster。

1.2 Super-P2P 网络

Super-P2P 网络模型的拓扑结构如图 1 所示。各 peer 节点在本地 Super-peer 的控制下直接建立连接,Super-peer 一般由几台 peer 节点互联构成,主要具有 3 个职能:(1)辅助本地

对等节点之间建立连接,通过集中认证,建立索引机制;(2)同时作为本地的服务提供者和服务接受者,参与通信;(3)与其他 Super-peer 节点构成对等结构,负责本地 peer 节点与其他区域 Super-peer 节点或 peer 节点的通信。

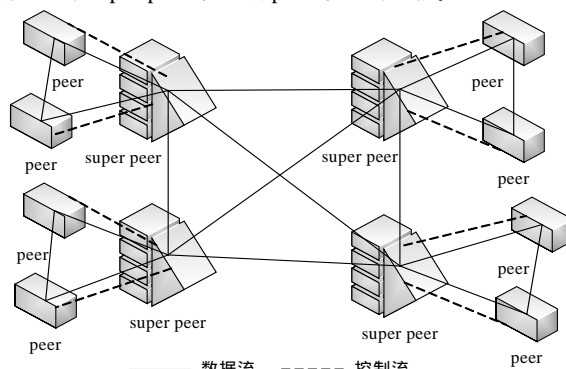


图1 Super-P2P 拓扑结构

Super-P2P 网络模型具有如下优点:

- (1)相对纯分散式 P2P 网络,具有更强的可控性,网络节点易于发现,同时 Super-peer 节点提供了安全服务;
- (2)相对于混合 P2P 网络,具有更强的容错性,由于

基金项目:重庆市科委自然科学基金资助项目(2007BB2192);浦东新区科技发展基金资助项目(PKK2005-07);国家发改委科学研究计划基金资助项目(2005-2137)

作者简介:李志国(1977-),男,博士,主研方向:知识管理,知识发现;冯永、钟将,博士;叶春晓,副教授

收稿日期:2007-09-07 **E-mail:** fengyong@cqu.edu.cn

Super-peer 节点是由几台 peer 节点构成的，只要不是全部节点失效，Super-peer 节点仍然能够提供服务；

(3)具有良好的可扩展性，只要适当增加 Super-peer 的规模，整个 Super-P2P 网络的规模就都可扩充。

2 基于 Super-P2P 结构的语义网络

现有的大多数网格^[7]都是围绕为终端用户提供计算服务而实现的，缺少解决领域问题的服务和知识服务。而语义网的研究给网格研究提供了一个契机，使网格从单纯的计算网络向可提供领域服务和知识服务的语义网络^[8]方向发展。

在语义网络的研究中，如何超越集中式的语义存储，实现更高层次的资源共享，是目前语义网络的研究热点之一。

2.1 基于 Super-P2P 的语义网络结构

基于 Super-P2P 结构的语义网络拓扑结构如图 2 所示。

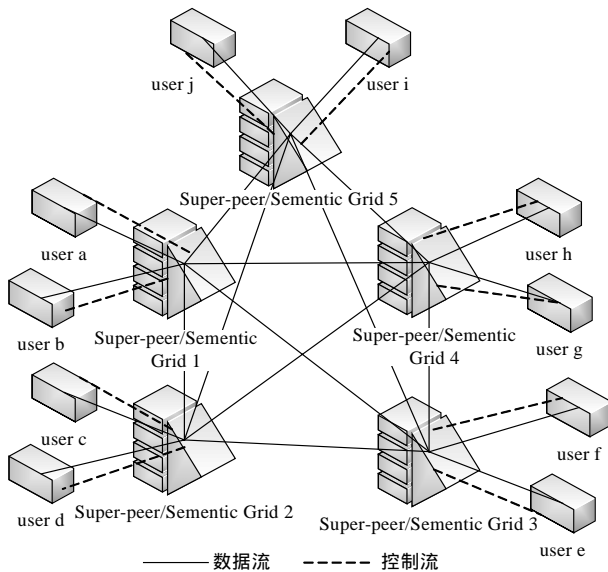


图 2 基于 Super-P2P 的语义网络拓扑结构

如图 2 所示，基于 Super-P2P 结构的语义网络是由多个语义网络互联构成的，增加或删除语义网络节点可以扩大或缩小整个网络结构，具有良好的可伸缩性。每个语义网络节点主要扮演如下角色：(1)整个结构中的一个 Super-peer 节点；(2)本地用户的服务提供者；(3)其他地区 Super-peer 节点的服务提供者；(4)其他地区 Super-peer 节点的服务接受者；(5)对本地用户节点有一定的控制作用，如增加、删除用户节点；(6)其他地区 Super-peer 节点的候选者。

相对于一般的 Super-P2P 结构，基于 Super-P2P 的语义网络结构具有如下区别：(1)本地一般用户不再作为 peer 节点，只接受来自本地语义网络的服务。这样减轻了一般用户的负担，增强了一般用户的安全性。(2)语义网络构成 P2P 结构，互为服务提供者和服务接受者。由于网络节点具有强大的计算服务功能、良好的容错性和可靠性以及增强的安全服务措施，因此整个 Super-P2P 语义网络能够更可靠、更安全地提供优质服务。(3)一旦本地语义网络失效，那么距离最近的语义网络充当本地的服务提供者，进一步保证了整个语义网络提供服务的能力。

2.2 基于 Super-P2P 的语义网络的资源发现机制

已有的研究成果表明利用小世界现象构造算法可以提高 P2P 系统的搜索效率^[9]。文献^[10]提出的 WS 模型^[10]是一种常用的 small-world 模型： n 个节点分布在一个圆环上，初始状态时，每个节点有 k 个连接，分别连向最近的 k 个点。然后，依

次调整各节点的连接，以概率 P 随机地改变连接的终端，但避免连向节点本身。

记 $D(i, j)$ 为节点 i 和节点 j 之间的最短距离，平均路径距离 L 的计算公式为

$$L = \frac{1}{n(n-1)/2} \sum_{i,j} D(i, j)$$

当 $P \approx 0$ 时， $L \rightarrow n/2k$ ，此时网络拓扑呈规则状态；当 $0.001 < P < 0.01$ 时， $L \rightarrow \ln n / \ln k$ ，此时节点不仅与相邻节点存在连接，还与远距离节点建立了少数的快捷连接，正是这些快捷连接有效缩短了 L ，使整个网络呈现出 small-world 特征。当搜索消息沿小世界环路传递时，其平均路径长度为 $O(n)$ 。

利用 small-world 的思想在 Super-P2P 语义网络节点进行资源搜索的过程描述如下：

- (1)请求节点根据要搜索的资源建立资源标识 R_ID ，检索本地语义网络 A 的资源库，如果匹配则 A 返回结果；否则 A 到路由表寻找距离其最近的直接连接节点 B ；
- (2) B 检索本地的资源库，如果有匹配项，则返回搜索结果，否则 B 返回本地搜索失败信息；
- (3) A 同时向其他所有直接连接节点发送搜索请求，若搜索成功，则返回结果；
- (4)否则 A 向其他所有已建立快捷连接的远程节点发送搜索请求，若搜索成功，则返回结果；否则，向用户返回搜索失败信息。

搜索原理示意图如图 3 所示。

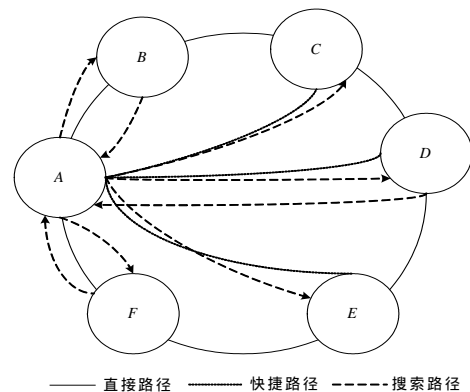


图 3 基于 small-world 的 Super-P2P 语义网络资源搜索

如图 3 所示， A 节点首先向距离最近的直接相连节点 B 发出搜索请求， B 返回失败信息； A 继续向直接相连节点 F 发出搜索请求， F 返回失败信息； A 同时向其已建立快捷连接的节点 C, D 和 E 发出搜索请求，最后 D 返回成功结果。

3 基于 Super-P2P 语义网络的知识服务

基于 Super-P2P 语义网络的知识服务体系如图 4 所示。服务被分成 3 个层次：

- (1)存储层：在本地语义网络节点中存储领域知识和其他语义网络节点的知识索引。
- (2)服务支持层：提供知识访问服务、知识表示服务、语义映射服务以及资源发现与匹配服务。整个知识服务层运行在 XML + RDF + Ontology 之上，XML + RDF + Ontology 构成网络结点之间以及语义网络和用户之间相互理解的基础。
- (3)用户层：与服务支持层通过基于 OWL 的查询和操作消息来进行通信，它从用户端接收知识查询请求，将其表示为 OWL 消息模式，然后通过通用网络环境发送消息给语义

网格知识服务处理器。

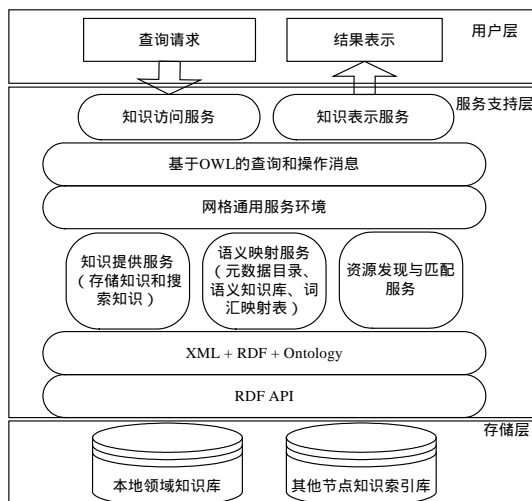


图4 基于 Super-P2P 语义网格的知识服务体系

3.1 知识访问与表示服务

用户使用知识访问服务接口向本地语义网格发送请求，并使用知识表示服务显示从本地语义网格返回的结果。知识访问与表示服务是语义网格的高层服务，可以单独部署在语义网格上。

3.2 知识提供服务

知识提供服务从知识访问服务中获取请求，然后转换为 OWL 格式，并在本地语义网格中搜索本地存储的知识，将得到的结果返回给用户。用户调用知识表示服务显示知识。

3.3 语义映射服务

将不同的局部本体映射为全局本体非常重要，这个服务被称为语义映射服务，它在背景知识和以下组件的基础上使用本体映射引擎执行映射服务^[11]。

(1)元数据目录：包括一些信息源的描述，如语义、存储路径、类型和供应者等。

(2)语义知识库：包括理解本体概念和属性所需要的知识，如同义词、中英文比较等。

(2)词汇映射表：为了获取一致的属性含义和数据描述视图，有必要定义一个全局词汇和本地 RDF 的相同语义描述表。

3.4 资源发现与匹配服务

当用户向语义网格节点发送请求时，资源发现与匹配服务利用 2.2 节描述的方法进行资源发现与匹配。

4 应用案例

中国教育科研网格(ChinaGrid)项目是“十五”211 工程公共服务体系“CERNET 高速地区网和重点学科信息服务体系建设”项目中的重要建设内容，依托 CERNET 建立可聚合与共享资源的公共服务平台——中国教育科研网格，对于实现信息资源的全面共享，提高教育信息化基础设施服务水平和高校教学科研水平，都将具有极其重要的意义。

在国家大力发展教育科研网格的背景下，利用本文提出的语义网格结构，建立面向多地区教育科研资源共享的教育科研网格具有重要的现实意义和广阔的发展前景。图 5 是基于 Super-P2P 语义网格的多地区教育科研知识共享平台。

平台构建方法如下：(1)根据本地区大学和科研机构的数量，建设本地区的语义网格节点；(2)每个语义网格节点提供

有本地区特色的知识服务，知识服务体系按照第 3 节的方式构建；(3)多个语义网格节点互连为 Super-P2P 结构，按照 2.2 节提供的方法进行资源发现；(4)整个结构可以扩充，高校和科研机构进行申请，获准后可以加入平台；(5)每个语义网格节点除了提供知识服务外，还应为本地区用户提供安全服务。

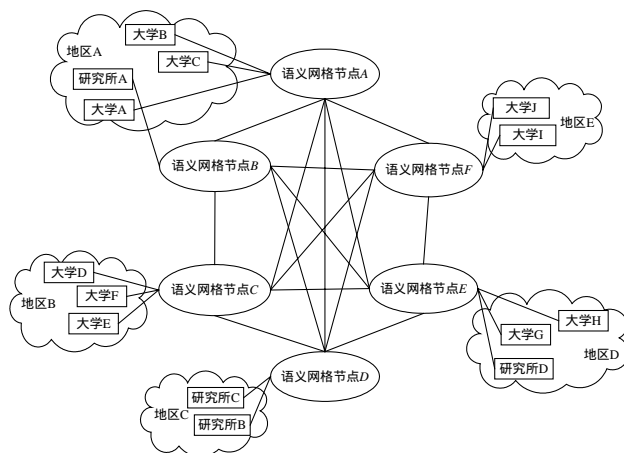


图5 基于 Super-P2P 语义网格的多地区教育科研知识共享平台

5 结束语

本文将 P2P 技术和语义网格技术相结合，提出一种基于 Super-P2P 语义网格的知识共享模型，以解决语义网格集中式语义存储的瓶颈，并提供分布式语义网格环境下的知识共享服务。通过应用案例分析，基于 Super-P2P 语义网格的知识共享模型，可以实现计算资源、存储资源、数据资源、信息资源、专家资源等多种信息资源的全面共享，提供高质量、高可靠性的知识共享服务。

参考文献

- [1] Berners-Lee T, Hendler J, Lassila O. The Semantic Web[J]. Scientific American, 2001, 284(5): 34-43.
- [2] Foster I, Kesselman C. The Grid: Blueprint for a New Computing Infrastructure[M]. San Francisco, CA: Morgan Kaufmann, 1999.
- [3] Foster I, Kesselman C, Tuecke S. The Anatomy of the Grid: Enabling Scalable Virtual Organizations[J]. International Journal of Super-computer Applications, 2001, 15(3): 200-222.
- [4] Roure D D, Hendler J A. E-Science: The Grid and the Semantic Web[J]. IEEE Intelligent Systems, 2004, 19(1): 65-71.
- [5] Dreamtech Software Group. Peer to Peer Application Development: Cracking the Code[M]. [S. l.]: Wiley Publishing, 2001.
- [6] Fiorano Software. Whitepaper: Super-peer Architectures for Distributed Computing[Z]. (2001-10-12). http://www.fiorano.com/white_papers/superpeer.pdf.
- [7] Foster I, Kesselman C. The Grid2: Blueprint for a New Computing Infrastructure[M]. San Francisco: Morgan Kaufmann, 2004.
- [8] Mario C, Domenico T. Semantics and Knowledge Grids: Building the Next-generation Grid[J]. IEEE Intelligent Systems, 2004, 19(1): 56-63.
- [9] 黄道颖, 黄建华, 庄雷, 等. 基于主动网络的分布式 P2P 网络模型[J]. 软件学报, 2004, 7(15): 1081-1089.
- [10] Watts D J, Strogatz S H. Collective Dynamics of Small-world Networks[J]. Nature, 1998, 393(6): 440-442.
- [11] 黄屹, 顾进广, 陈华萌, 等. 网格环境下基于本体的知识库模型研究[J]. 武汉大学学报(理学版), 2005, 51(5): 603-608.