

# 集成语义和多 Agent 的网格资源发现模型

熊曾刚<sup>1,2</sup>, 杨扬<sup>1</sup>, 刘丽<sup>1</sup>, 陈福<sup>1</sup>, 曾明<sup>1</sup>

(1. 北京科技大学信息工程学院, 北京 100083; 2. 湖北孝感学院计算机科学系, 孝感 432000)

**摘要:** 将P2P技术应用到网格环境中, 可为网格的可扩展性提供更好的解决方案。该文提出一种基于P2P模式的网格资源发现模型, 将资源的发现分为2个阶段: 网格节点GridPeer域内查询和网格节点GridPeer域间查询, 并且集成了语义和多Agent技术。理论分析表明, 该模型比集中式资源发现方案可扩展性更强, 更适用于基于P2P模式的网格环境下资源的匹配。

**关键词:** 网格; 对等网; 语义; 多agent; 资源发现

## Grid Resource Discovery Model with Semantic and Multi-Agent

XIONG Zeng-gang<sup>1,2</sup>, YANG Yang<sup>1</sup>, LIU Li<sup>1</sup>, CHEN Fu<sup>1</sup>, ZENG Ming<sup>1</sup>

(1. School of Information Engineering, University of Science and Technology Beijing, Beijing 100083;

2. Department of Compute Science, Hubei Xiaogan University, Xiaogan 432000)

**【Abstract】** The integration of grid and P2P provides a solution for the scalability of conventional grid. This paper proposes a P2P-based grid resource discovery model, which is integrated semantic and multi-Agent, to optimize grid resource discovery. In this model, resource discovery consists of two phases: intra-GridPeer query and inter-GridPeer query. Theory analysis proves that the model is more flexible and more scalable, which is more suitable for P2P Grid resources matchmaking.

**【Key words】** grid; P2P; semantic; multi-agent; resource discovery

### 1 概述

资源发现机制是网格环境下面临的最重要的挑战之一。传统的网格资源主要是由集中式和分层式模型来进行管理的。资源发现仍然存在潜在的性能问题和安全瓶颈以及单点失效问题<sup>[1]</sup>。

最近的研究显示, 网格和P2P可以很好地结合, 从而更好地发现资源<sup>[2]</sup>。在大规模动态异构网格环境下使用语义来精确发现资源也是一个相当具有挑战性的主题。虽然目前基于语义的网格资源匹配方法已被提出<sup>[3]</sup>, 但是这些方法并不完全适合分布式网格环境。

本文提出一个基于P2P模式的网格资源发现模型, 该模型集成了语义和多Agent技术, 克服了传统的集中式资源发现机制的弊端, 用语义的方法进行资源匹配, 能够增强资源匹配的准确性, 优化网格资源发现。

### 2 集成语义和多 Agent 的网格资源发现模型

#### 2.1 基于P2P模式的网格模型

本文提出的基于P2P模式的网格模型借鉴了文献[4]的思想。如图1所示, 模型的底层采用了传统的网格技术, 在GridPeers之间采用了P2P模式来进行通信。这里, 把GridPeer的管理域称为网格社区。

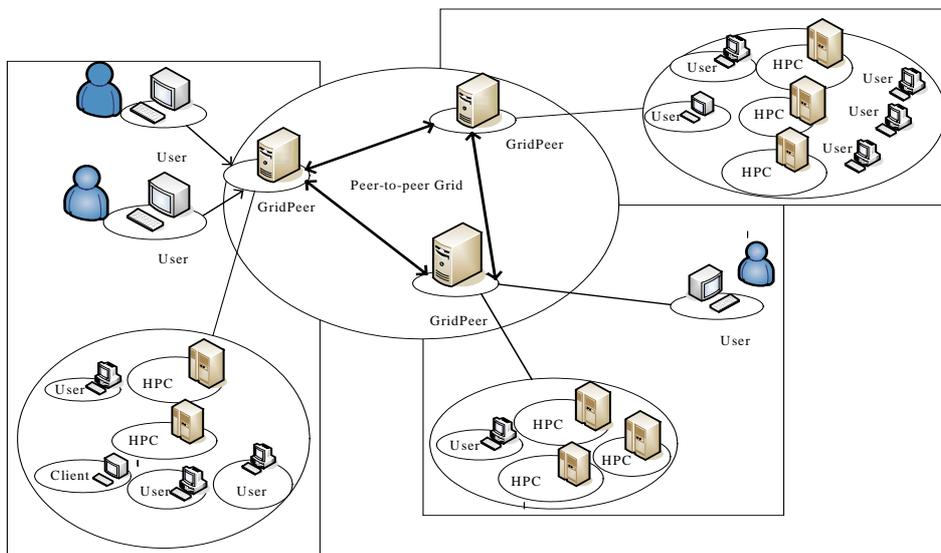


图1 基于P2P模式的网格模型

#### 2.2 GridPeer 内部框架

文献[5]指出, 网格与Agent能够聚合在一起。一方面网格能够改进Agent系统的鲁棒性结构; 另一方面, Agent能

**基金项目:** 国家自然科学基金资助重大项目(90412012); 国家自然科学基金资助项目(60673160); 孝感学院自然科学基金资助项目(Z2007039)

**作者简介:** 熊曾刚(1974-), 男, 讲师、博士研究生, 主研方向: 网格计算, workflow; 杨扬, 教授、博士生导师; 刘丽, 副教授、博士; 陈福、曾明, 博士研究生

**收稿日期:** 2007-05-25 **E-mail:** jkxxzg@163.com

够增加网络的自治性和灵活性。图 2 显示了 GridPeer 的基本结构。在 P2P 网络社区, GridPeer 扮演着超级节点的角色, 同时它由 2 种 Agents 组成: RequestAgent 和 BrokerAgent。

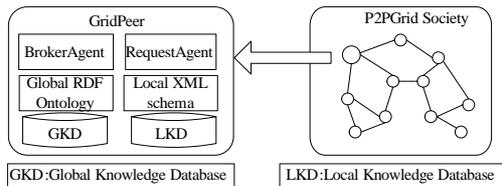


图 2 GridPeer 内部结构

RequestAgent 负责 GridPeer 内部的查询服务请求。当服务请求调用网格服务时, 向 GridPeer 的 RequestAgent 发送请求信息, RequestAgent 在本地知识库(Local Knowledge Database, LKD)中查找对应的网格服务。如果本地知识库中存在对应的网格服务, RequestAgent 就返回一个信息; 否则, 将请求信息转发给 BrokerAgent。

BrokerAgent 完成 2 个任务: 将本地 XML schema 映射成全局 RDF 本体和将 RequestAgent 的信息转发到另一个 GridPeer。一个 BrokerAgent 知道存储在 GridPeer 间的所有本体信息, 而且通常与其他 GridPeers 的本体相关联。

### 3 基于 P2P 模式的网格资源发现算法

本体为基于 P2P 模式的网格节点 GridPeer 域提供了一种很好的资源发现方法, 如图 2 所示的结构。由于 XML 已经成为互联网上信息描述和信息交换的标准, 因此在 GridPeer 域内采用 XML 数据库形式进行存储, 在 GridPeer 域间采用本体形式进行存储。其中, RDF(Resource Description Framework)是目前广泛使用的描述本体的建模语言, 它本身包含了类、属性、域约束、子类以及子属性关系等。RDF 的语法采用 XML 语法形式, RDF 的目标之一是基于 XML 以标准和交互方式来尽可能地规范数据语义。

本文采用概念模式来解决 XML 源数据中结构多样性的问题。

#### 3.1 转换本地 XML schema 到全局 RDF 本体

一个主机无论什么时间加入到基于 P2P 模式的网格环境中, 主机通过从本地 XML schema 到全局本体建立映射来进行注册和索引。这个映射过程通过 GridPeer 中 BrokerAgent 的全局知识库(Global Knowledge Database, GKD)来实现 XML schema 的匹配和存储。XML schema 匹配过程中的关键问题是 XML 源文档结构和 RDF 源域结构的保存。将本地 XML schema 映射到全局本体中的转换过程是通过 GridPeer 中 BrokerAgent 来实现的。BrokerAgent 将 XML Schema 的属性(attribute)和简单元素(simple element)转换成 RDF 的属性; 将 XML schema 复杂类型元素(complex element)转换为 RDF 的类(class)。同时将 XML schema 中的 element-attribute 关系和 element-subelement 关系分别转换为 RDF 中的 class-to-literal 关系和 class-to-class 关系。定义一个新的 RDFS 谓词 rdfs:contain 来表示 class-to-class 的关系, 这里, rdfs 表示了一个新的命名空间 <http://sales.org/rdfs#>。在它的域内增加一个新的属性作为一个类(此类由其父元素转换而来), 它的值域区间作为另一个类(此类由子元素转换而来), 名字用 rdfs:contain 来表示。图 3 列出了一个具体的 XML schema A 与全局 RDF 本体 B 的转换关系。在图 3 中, 全局 RDF 本体 B 中 rdfs:contain 能够表示嵌套关系。例如, 在全局 RDF 本体 B 中谓词 rdfs:contain 表示 CD 音乐的作曲者, 可以在 XML

schema A 中通过 XPath/Compact Discs/CD/Composer 来表示。

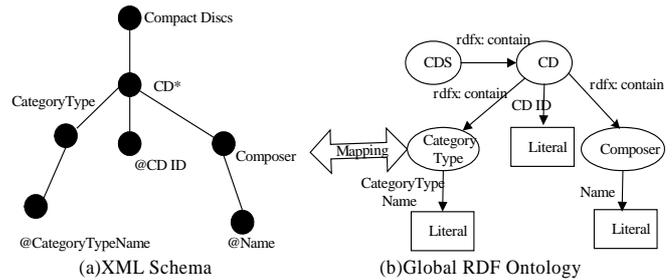


图 3 本地 XML schema A 与全局 RDF 本体的映射

表 1 列出了图 3 所示的在全局 RDF 本体 B 与 XML schema A 之间的映射关系。

表 1 全局 RDF 本体 B 与 XML schema A 之间的映射关系

XPath of XML schema A	Global RDF ontology B
/ Compact Discs	Compact Discs (CDS)
/Compact Discs/CD	CD
/Compact Discs/CD/CategoryType	CategoryType
/Compact Discs/CD	CategoryType.
/CategoryType/@ CategoryTypeName	CategoryTypeName
/Compact Discs/CD/@CDID	CD.CDID
/Compact Discs/CD/Composer	Composer
/Compact Discs/CD/Composer/@Name	Composer.Name

#### 3.2 网格环境下集成语义和多 Agent 的资源发现算法

基于 P2P 模式的网格环境下资源发现主要由 2 部分组成: (1)满足条件的资源属于一个或者多个外部的 GridPeer; (2)在 GridPeer 内, 即本地网格社区内的本地资源请求可以通过 GridPeer 直接返回给主机。

为了能够表示资源查询过程, 假定在 GridPeer 内 XML 查询遵循 XQuery 规范。用类似 SQL 语法的 RDQL 语言来表示 GridPeer 内的资源查询。为了简化操作, 将 XQuery 查询  $Q$  用三元组  $(V_{Q^s}, V_{Q^w}, C_Q)$  来表示, 其中,  $V_{Q^s}$  和  $V_{Q^w}$  分别表示包含所有 XML 路径的返回结果子集和 where 条件子集;  $C_Q$  包含了查询的约束条件, 其表达形式是  $vRc$ , 这里  $v \in V_{Q^w}$ ,  $C$  代表一个常量;  $R$  代表一个比较参数(如=, <, >, <=, >=, ≠)。同样, 也可以用一个三元组  $(P_{Q^s}, P_{Q^w}, C_Q)$  来表示 RDQL 查询  $Q$ 。其中,  $P_{Q^s}$  和  $P_{Q^w}$  分别包含了所有 RDF 路径的选择条件和 where 条件的子集。  $C_Q$  代表约束条件, 其表示形式为  $pRc$ 。其中,  $p \in P_{Q^w}$ ,  $C$  代表一个常数;  $R$  代表一个比较参数。

基于 P2P 模式的网格环境下集成语义和多 Agent 的算法过程如下:

(1)主机通过 XQuery 方式向 GridPeer 发送资源请求。一旦 GridPeer 接收到一个 XML 查询, RequestAgent 在本地 LDK 中查找是否有满足条件的资源。如果有, RequestAgent 将查询结果返回。否则, 将 GridPeer 将查询请求转发到 BrokerAgent, 转到步骤(2)。

(2)BrokerAgent 使用 3.1 节的方法将本地 XML schema 转换成全局 RDF 本体, 将 XQuery 查询条件转换成 RDQL 查询条件。然后转到步骤(3)。

(3)BrokerAgent 通过洪泛法或者随机漫步法将 RDF 的查询条件转发给其最近的 GridPeers 节点, 如果没有发现满足条件的资源, 则返回错误信息。否则, 转到步骤(4)。

(4)比较最相近的 GridPeers, 将 RDF 查询条件与 BrokerAgent 中 GKD 中的全局 RDF 本体进行比较, 如果找到满足条件的资源, 再利用 3.1 节的方法转换本地 XML schema 到

全局 RDF 本体, 将全局 RDF 本体转换成 XML schema, 找到 GridPeer 下的资源提供者。然后将结果返回给对应的消费者。

### 3.3 应用实例

根据图 3 所示的内容举 2 个实例, 以此说明本地 XML schema 与全局 RDF 本体的转换规则。

**例 1** 如果有一个查询: “找到 CDS 的作曲者 jazz”, 用户首先按照图 3 所示的规则向 GridPeer 提交一个本地 XML schema 查询语句。本地 XML 源 A 与查询结果 Q 如下:

```
The target XML query is :
For $CD in doc("CDS.xml")/CDS/CD
  Where $ /CD/Composer/@Name=" jazz"
  Return $CD/@CDID
For the local XML source A:
VQr ={/Compact Discs/CD/@CDID }
VQw ={/Compact Discs/CD/Composer, /Compact Discs/CD/Composer/@Name }
CQ ={/Compact Discs/CD/Composer/@Name ,=" jazz"}
```

**例 2** 如果在 GridPeer 内没有发现 CDS 的作曲者 jazz, BrokerAgent 发送一个全局请求本体信息(前缀 go 代表命名空间 “http://sales.org/global#”, 在此命名空间定义全局本体)。查询结果 Q 显示如下:

```
SELECT ?CDID
WHERE (?Composer, <go: Name>, ?Name)
      (?CD, <go: CDID>, ?CDID)
AND (?Name, eq, "jazz")
PQr = {CD.CDID}
PQw = {Composer, Composer.Name, CD, CD.CDID}
CQ = {(Composer.Name, eq, "jazz")}
```

(上接第 6 页)

言, 属于相互独立的并发事件, 且均不需移动链表元素, 因此其时间复杂度  $T(n)$  和链表中的元素个数  $n$  无关, 即  $T(n)=O(1)$ 。

基于车站索引的列车信息查询, 其索引表、主表中的数据呈非递减有序排列, 工程常用顺序查找算法或二分查找算法。一般用平均查找长度(ASL)评价查找算法的效率。以图 5(b) 为例, 索引表的长度为  $m+1$ , 主表中对象的个数为  $3 \times m+1$ , 以区间双向链表为子表并设其长度为  $s$ , 可以对上述两种查找算法的效率进行评价, 见表 1。

表 1 2 种查找算法的效率评价

算法名称	ASL
顺序查找算法	$\frac{m+2}{2} + \frac{s+1}{2}$
二分查找算法	$\frac{m+2}{m+1} \lg(m+2) + \frac{s+1}{2} - 1$

显然, 二分查找算法的效率优于顺序查找算法, 因此, 部、局宜采取二分查找算法实现列车信息的查询。另一方面, 受列车追踪运行安全间隔的限制, 区间追踪运行的列车数量  $s$  一般不会很大, 索引表的长度为  $m+1$ , 即线路上车站数量愈多, 索引表愈长, 它对查找效率影响较大, 合理的索引表长度往往是至关重要的。而索引表长度的减少以其数量增加为代价, 宜在二者之间取得平衡。

## 4 结束语

本文提出了一种基于 P2P 模式的网格环境下集成语义和多 Agent 的资源发现模型。本系统能够在 GridPeers 间通过本体概念对 schema 编码以支持在非结构 Overlay 层实现语义匹配的请求。描述了在该模型下如何支持在 GridPeer 内部和 GridPeers 之间实现语义请求的服务查询。用户查询能够准确地异构 XML 和 RDF 数据之间进行传送, 因此, 在 P2Pgrid 中, 资源的访问对用户是透明的。下一步的工作是在此模型基础之上开发一套原型系统, 比较本文模型与其他基于语义的模型下资源发现的性能。

### 参考文献

- [1] Foster I, Iamnitchi A. On Death, Taxes, and the Convergence of Peer-to-peer and Grid Computing[C]//Proc. of the 2nd Int'l Workshop on Peer-to-peer Systems. Berkeley, USA: [s. n], 2003: 118-128.
- [2] Iamnitchi A, Talia D. P2P Computing and Interaction with Grids[J]. Future Generation Computer Systems, 2005, 21(3): 331-332.
- [3] Tangmunarunkit H, Decker S, Kesselman C. Ontology-based Resource Matching in the Grid—the Grid Meets the Semantic Web[C]//Proceedings of SemPGRID'03. Budapest, Hungary: [s. n], 2003.
- [4] Yang B, Garcia-Molina H. Designing a Super-peer Network[C]//Proceedings of International Conference on Data Engineering'03. Bangalore, India: [s. n], 2003: 49-60.
- [5] Foster I, Jennings N R, Kesselman C. Brain Meets Brawn: Why Grid and Agents Need Each Other[C]//Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multiagent Systems. New York, USA: [s. n], 2004-07.

## 5 结束语

移动闭塞系统作为一种先进的自动闭塞系统, 符合我国铁路扩能提效的发展需求。本文针对铁路运输特点, 就增强移动闭塞系统列车运行控制与安全防护功能和充分发挥移动闭塞的积极作用, 对铁路移动闭塞系统中列车信息的移动性管理进行了探讨, 还需要对下列问题继续进行深入研究:

(1) 层次索引表的跳跃查询策略研究。跳跃表(Skiplist)是一种高效数据结构<sup>[5]</sup>, 根据铁路运输网络的拓扑结构和车站规模、等级等因素, 合理选择若干车站作为跳跃平台, 进一步加快查找速度。

(2) 列车信息及移动性管理信息的可靠传递。列车、车站、控制中心之间实现信息的可靠传递, 才能实现列车追踪运行的安全控制与防护。

(3) 结合地图匹配、虚拟现实技术, 利用列车精确定位信息实现移动闭塞条件下列车追踪运行虚拟场景的实现。

### 参考文献

- [1] 谢肇桐. 移动闭塞系统[J]. 铁道通信信号, 1996, 32(2): 35-36.
- [2] 张 勇. 新型列控系统-移动自动闭塞条件下线路通过能力的理论分析及计算机仿真[D]. 北京: 北方交通大学, 1998: 22-25.
- [3] 宁 滨. 轨道交通系统中的列车运行追踪模型及交通流特性研究[D]. 北京: 北方交通大学, 2005: 11-32.
- [4] 朱战立. 数据结构——使用 C++ 语言[M]. 西安: 西安电子科技大学, 2001: 63-73.
- [5] 王晓东. 计算机算法设计与分析[M]. 北京: 电子工业出版社, 2001: 210-216.