

# 教育资源网格中资源的描述与组织技术

马 艳, 刘方爱

(山东师范大学 信息科学与工程学院, 山东 济南 250014)

**摘要:** 网格中的资源发现是网格研究的重要问题。文中主要针对教育资源网格中的资源发现, 而教育资源网格中资源的描述和组织技术是资源发现的基础。通过分析国家《基础教育教学资源元数据规范》, 采用资源描述框架来描述教育资源和资源请求, 用索引链表的存储方式来组织教育资源。通过与其他结构的比较分析, 证明了索引链表结构能减少查询次数, 适应网格环境的动态性和异构性, 更有利于资源发现。

**关键词:** 教育资源; 资源组织; 资源描述框架

**中图分类号:** TP393.01

**文献标识码:** A

**文章编号:** 1673-629X(2008)04-0059-04

## Resource's Description and Organization Technology in Education Resource Grid

MA Yan, LIU Fang-ai

(School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China)

**Abstract:** Grid resource discovery is a most important problem in grid research. Mainly researches on education resource grid, and resources' description and organization technology are the foundation of efficiently discovering resources. By analyzing metadata specification of China base on education resource, it uses resource description framework to describe education resource and resource requests, presents index and link table to organize education resource. Based on comparison of other structures, it proves that index and link table can reduce the time of request, adapt dynamic and heterogeneous characteristics in grid environment and is advantageous to the resource discovery.

**Key words:** education resource; resource organization; RDF

### 0 引言

网格(grid)技术用于实现广泛的分布资源共享和协同工作。在很多场合, 资源需要分布存放、实施分布管理的方法和技术。教育资源网格正是针对中小学的教育资源共享问题, 研究分布式教育资源的管理策略, 实现教育资源网格平台。教育资源网格平台提供一种无缝的、集成的资源共享和协作环境, 其目标是协调网格中资源的使用, 以便及时响应网格用户的资源请求。但是网格中潜在着巨大数量的资源和用户以及资源具有异构性、动态性和自治性等特点, 如何使网格应用程序方便地使用各种资源便成了必须要解决的问题。而使用一种适合于资源特点的资源描述和组织技术便成了定位使用资源的前提。

资源信息的组织需要遵循:

(1) 资源分类应具有多个层次, 同时组织结构要体现不同类型资源的差异;

(2) 资源的属性可以冗余, 保证资源的完整表示。文中所作的工作:

一是在教育资源网格体系结构的基础上, 根据教育资源的特点和《基础教育教学资源元数据规范》, 用RDF(Resource Description Framework)元数据来描述网格中的教育资源和查询请求;

二是借鉴MDS的经验, 对教育资源进行分类, 用索引链表结构来组织教育资源;

三是比较分析表明采用上述教育资源的组织方式, 可以剪枝查询空间, 提高查询效率。

## 1 教育资源网格体系结构

### 1.1 系统架构

教育资源网格采用扩展树结构, 如图1所示。

逻辑结构自顶向下形成一个层次式的树形结构。树的根节点为一个区域中心节点; 树的叶节点表示网格的最低层, 即各个学校, 它们只需要维护自身的资

收稿日期: 2007-07-30

基金项目: 国家自然科学基金(90612003)

作者简介: 马 艳(1982-), 女, 山东泰安人, 硕士研究生, 研究方向为网格技术、网络安全; 刘方爱, 教授, 博士, 博士生导师, 研究方向为并行/分布处理、并行算法。

源;中间节点可以是省、市、县教育管理部门。兄弟节点之间建立了一个环形结构,目的是为了提系统的容错性和保证负载均衡。注意,具体的资源分散在各自的学校,而资源的属性信息存放在树的各个中间节点上,以供应用程序查询。

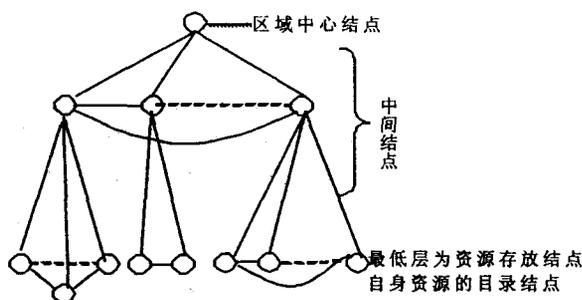


图 1 教育资源网格体系结构

### 1.2 架构中各层节点所完成的功能

#### (1)区域中心节点。

中心节点维护整个系统的目录信息,以便用户对不同子域内资源的查询与访问;中心节点是属于全局统一管理的,通过分析各个地区资源的访问量,找出热点资源,提高查询的成功率,还可调整价格和负载均衡;中心节点还负责响应中间节点的请求、本系统内的成员认证、计费管理等相关问题。

#### (2)中间节点。

中间节点是指省、市、县教育管理部门,它维护其辖域内的目录信息,即维护本域内各学校目录节点的地理信息,当学校节点的资源发生变化时,会实时地更新到中间节点;中间节点负责响应低层资源节点的请求。

#### (3)最低层资源节点。

最低层资源节点是指学校节点,一个学校拥有多个教育资源,学校的所有资源信息存储在学校自己的节点上。该类节点管理本地资源,包括添加、删除、修改,实时地向上层节点报送本地资源的动态变化;维护本地资源的目录信息;响应本地用户对资源的查询请求。

## 2 教育资源和资源请求的描述

### 2.1 RDF

网格资源描述技术有多种,如 RDF、RSL(Resource Specification Language)、WSDL(Web Service Description Language)等<sup>[1]</sup>。通过分析 RSL、WSDL 和 RDF 各自的特点和局限性,提出了资源描述框架 RDF 能更准确地描述网格资源<sup>[2]</sup>。也采用了 RDF 来描述网格资源。用 RDF 描述资源,资源提供者可以给出资源的细节描述,而资源请求者可以自描述他们的请求,使查询更加

精确和灵活。

RDF 描述网格资源有很多优点<sup>[1]</sup>:通用性;智能和准确性;综合性。故这里采用 RDF 来描述教育资源和资源请求。

### 2.2 基础教育教学资源元数据规范

关于基于网络的教育资源的标准化,有许多教育元数据标准,最有影响力的是:

- 1) IMS 的 Learning Resource Metadata(学习资源元数据规范);
- 2) IEEE LTSC (Learning Technology Standards Committee,学习技术标准委员会)的 LOM(Learning Object Metadata,学习对象元数据);
- 3) OCLC(Online Computer Library Center)Dublin Core 的 Dublin Core 元数据标准。

我国也制定了“中国教育信息化技术标准体系”(CELTS)<sup>[3]</sup>。该体系包括:《学习对象元数据》、《教育资源建设技术规范》和《基础教育教学资源元数据规范》。《基础教育教学资源元数据规范》核心元数据有 23 个,其中必需元素包括标题、学科、关键词、描述、标识、格式、日期、语种、类型、作者、适用对象 11 个核心元素。在描述教育资源时对基础教育教学资源元数据规范作了少量修改。举例,某一资源的元数据描述如表 1 所示。

表 1 某资源的元数据表示

|      |  |                  |            |
|------|--|------------------|------------|
| 标题   | 牛顿第二定律   |                  |            |
| 学科   | 物理   |                  |            |
| 关键字  | 牛顿第二定律,加速度   |                  |            |
| 描述   | 这是一个关于牛顿第二定律的课件,在水平桌面上的物体在受到不同拉力作用时运动所需时间也不一样,即加速度也不一样,实验证明加速度大小与外力大小成正比 |                  |            |
| 作者   | 名称:张三  |                  |            |
|      | 单位:山东省实验中学   | 地址:济南市经七路 73 号   | Email: * 1 |
| 适用对象 | 高中一年级  |                  |            |
| 出版者  | 名称:山东省实验中学   |                  |            |
|      | 地址:济南市经七路 73 号   | 电话:0531-86926496 | Email: * 2 |
| 标识   | http://www.sdshiyen.cn/shiyen/111.ppt                                    |                  |            |
| 价格   | 10 元   |                  |            |
| 格式   | 范围:43mb  |                  |            |
| 日期   | 2006-11-15   |                  |            |
| 语种   | 汉语   |                  |            |
| 类型   | 视频类素材  |                  |            |

### 2.3 用 RDF 描述教育资源

一个 RDF 文件包含多个资源描述,而一个资源描述是由多个语句构成,每个语句又是由资源、属性类型、陈述构成的三元体。每个资源可以由 URI 或者加 Id 来命名。从表 1 可以得知该资源的精确属性信息,

用 RDF 描述出来,如图 2 所示。

```
<? xml version="1.0"? >
<rdf:RDF xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns# xmlns:s="http://description.org/schema/">
<rdf:Description about="http://www.sdshiyen.cn/shiyen/111.ppt">
<s:title>牛顿第二定律</s:title>
<s:subject>物理</s:subject>
<s:key>
<key1>牛顿第二定律</key1>
<key2>加速度</key2>
</s:key>
<s:description>是一个关于牛顿第二定律的课件,在水平桌面上的物体在受到不同拉力作用时运动所需时间也不一样,即加速度也不一样,实验证明加速度大小与外力大小成正比
</s:description>
<s:author>
<name>张三</name>
<unit>山东省实验中学</unit>
<address>济南市经七路 73 号</address>
<email>*1</email>
</s:author>
<s:app-object>高中一年级</s:app-object>
<s:publisher>
<name>山东省实验中学</name>
<address>济南市经七路 73 号</address>
<tel>0531-86926496</tel>
<email>*2</email>
</s:publisher>
<s:identifier>http://www.sdshiyen.cn/shiyen/111.ppt</s:identifier>
<s:price>10 元</s:price>
<s:format>范围:43mb</s:format>
<s:date>2006-11-15</s:date>
<s:language>汉语</s:language>
<s:kind>视频类素材</s:kind>
</rdf:Description>
</rdf:RDF>
```

图 2 某资源的 RDF 元数据的 XML 表示

特定资源与属性类型、属性值在一起构成 RDF 陈述。陈述可以用有向图来表示。图 3 中,椭圆框表示资源(或其他主体),弧表示属性,表示属性值的字符串被放在长方形框里。将表 1 用图解法表示,如图 3 所示(注:属性没有全部画出,其中张三和山东省实验中学又是具有多个属性的主体)。

另外,教育资源的请求也用 RIF 来描述。

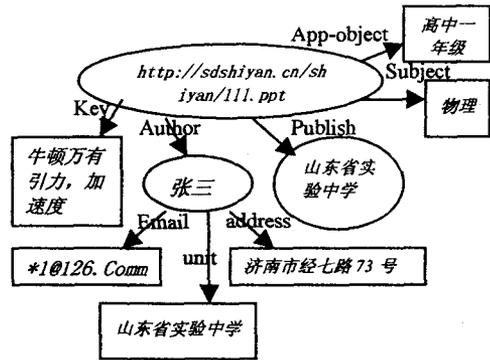


图 3 某资源的 RDF 图形表示

### 3 教育资源的组织方式

国家教育资源建设规范中规定:学科、适用对象和类型是资源分类的 3 个主要依据。这 3 种分类标准可以组合成 6 种不同的资源分类体系,方便资源库的资源组织和管理。按照科目-适用对象-类型的顺序来对教育资源进行分类,采用链表和索引表相结合的方法组织资源,具体如图 4 所示。

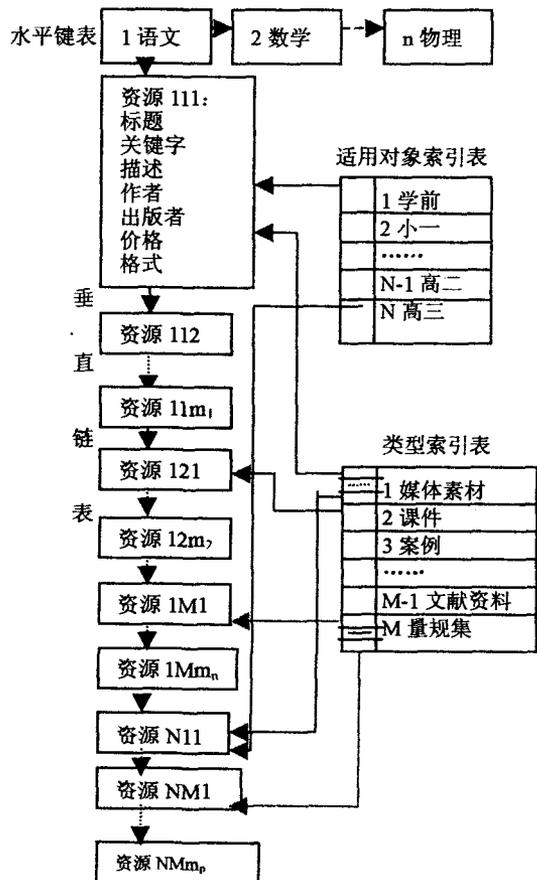


图 4 教育资源的组织方式图

该图由 3 部分组成:水平链表,垂直链表和索引表。

1) 水平链表是按照科目类别依次连接的链表。因为每个学校有最擅长的学科,故水平链表按照该学

校各门学科的教学质量优劣进行排序,排在水平链表前面的是该学校最擅长的学科,最擅长学科可以根据下载次数或教育局的教学评估结果等界定。

2) 索引表分为适用对象索引表和类型索引表。它们和水平链表中的科目共同决定教育资源的类别。确定了科目后,就直接从该科目下面的垂直链表查找资源信息,而不用搜索所有的资源信息,大大剪枝了查询空间。科目确定后,再从适用对象索引表和类型索引表找到用户请求中相应的适用对象和类型,这样就可以直接定位该类资源的位置。适用对象索引表中,每项有一个指针域。如“学前”指向语文-学前-媒体素材类的第一个资源;同理,“高三”指向语文-高三-媒体素材类的第一类资源,即该表中的每项指向该项出现的初始位置。类型索引表中,每项有  $N$  个指针域。如“媒体素材”的第一个指针指向语文-学前-媒体素材类的第一个资源,最后一个指针指向语文-高三-媒体素材类的第一个资源;同理,“量规集”的第一个指针指向语文-学前-量规集类的第一个资源,最后一个指针指向语文-高三-量规集类的第一个资源。

3) 垂直链表是某一科目的所有资源依次连接的链表。资源不是杂乱地连在一起,是有规律的。资源是按照类别连接的,比如图 4,语文下的垂直链表中,类别依次是:语文-学前-媒体素材,语文-学前-课件,……,语文-学前-量规集,……语文-高三-媒体素材,……语文-高三-量规集。同时每一类的资源按照热度(下载次数)进行排序,下载次数最多的资源排在该类资源的第一个,依次类推。初始建立网格时,同类资源的顺序可根据教学评估结果排序。其中,“语文”项下的“资源 111”中的三个数字分别代表:第一个数字表示适用对象是 1 学前,第二个数字表示类型是 1 媒体素材,第三个数字表示语文-学前-媒体素材类的第 1 个资源;同理,“语文”项的最后一个“资源  $NM_{m_p}$ ”则表示语文-高三-量规集类的第  $m_p$  个资源(即最后一个资源)。“数学”等其他学科下的垂直链表和索引表类似于“语文”学科。垂直链表中的每个节点存的是某资源的元数据信息,如标题、关键词、描述、作者和格式等。每个资源根据情况不同,元数据的个数可以不同。其他资源的属性描述与“资源 111”类似。请求发到某学校节点后,就用这些元数据信息与请求进行比较,满足要求的结果返回给用户。

举例:先考虑用户在本地请求资源,就可以满足要求的情况。图 4 假设是节点山东实验中学存储的资源信息,用户提出查询请求:科目-语文;适用对象-高一;类型-文本类素材;标题-《荷塘月色》教案示例;

关键字-教学案例,荷塘月色,《荷塘月色》教案示例;描述-荷塘月色的教案;格式-msword。首先科目是语文,在水平链表中找到语文,再沿着语文项的垂直链表向下搜索,搜索前先从适用对象索引表找到“高一”的索引项,再从类型索引表中找到“文本类素材”的索引项,就可以直接定位到‘语文-高一-文本类素材’类型资源的位置,然后将用户的请求与资源的元数据进行比较,将满足要求的结果按照热度显示出来,供用户二次选择。

#### 4 比较分析

资源信息的目录组织方式有很多种。

Globus MDS 实现了基于 LDAP 的树状元数据目录服务,实体描述为属性/值对。每个组织内通常都有一个集中的索引服务,组织规模比较大时,需要形成树状层次结构。但是资源的查询和访问是通过资源的唯一标识实现的,而不是通过它的属性。虽然目录信息树也包含一系列的属性/值对,但这只是它的内容描述,不是元数据描述,不能用来进行多属性的查询;同时对于大型的、动态刷新和复制频繁发生的网格系统,MDS 具有被动更新的缺点,不能很好地工作。

结构化的 P2P 中,采用了分布式哈希表的方法,将资源映射为一个 Hash 值,具有精确关键字映射的特性,实现了负载均衡。但也因为它支持精确关键字匹配,使得无法支持内容/语义等查询,也无法与信息检索领域的研究成果结合;而且分布式哈希表维护起来很复杂。虽然文献[4,5]将分布式哈希表扩展到了支持关键字和多属性查询,但是它导致负载和更新的开销极大。

文中借鉴 MDS 中根据资源分类的特性,对资源信息进行组织,同时遵循资源分类具有多个层次的原则。不同的是把 LDAP 中以一条条记录存储的方式变成了链表存储。顺序与链式存储方式的优缺点很明显:顺序利于直接存取,链式便于插入和删除。一般用户查询是根据科目、标题、适用对象、类型、关键字等属性信息进行查询,用户是不了解网格结构的,很少有用户会对资源直接定位存取,所以顺序存储方式的优点不能完全发挥;而虽然学校节点是稳定节点,但是学校上的资源是不稳定的,会随时离开/加入,所以学校节点上的资源信息要进行动态更新,就是要删除/插入垂直链表中的资源信息,很明显,链式存储结构较适合于动态更新的网格环境。同时,在链表的基础上引入了索引表,它帮助剪枝了查询空间,减少了查询成本,提高了查询效率。

(下转第 66 页)

- (3)  $Z = Y + D$ ;  
 (4)  $P = P(Z, Q(S_i, A_i))$ ;  
 (5) If  $p_c < 0$  then Add( $X, Y$ ) Else Add( $X, Z$ );  
 (6) If Full( $T$ ) then TrainBP( $T$ );  
 (7)  $S_i = S_{i+1}$ , 转步骤(1); 至学习结束。

这里  $X$  为状态 - 动作对( $s, a$ ),  $V_{BP}$  为神经网络输出,  $Q(S_i, A_i)$  为强化学习得到的  $Q$  值,  $Z, Y, D$  都经过了时延。

#### 4 实验结果及分析

文中的 RoboCup 客户端平台底层代码基于 UVA01, 服务器采用 sim9.3.7 版本。总共进行 1000 个训练周期, 每个训练周期为 5 个服务器周期。实验结果如图 4 所示。

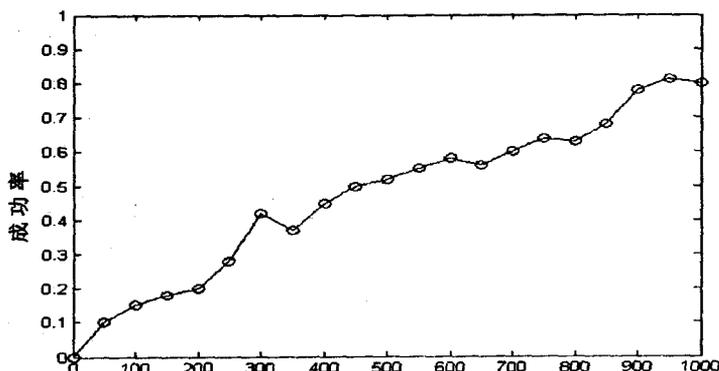


图 4 传球成功率统计曲线

从仿真实验结果来看, 传球的成功率相比以往有了较大的提高, 基于神经网络的强化学习算法在一定程度上提高了传球的成功率。用神经网络来存储  $Q$  值函数与常见的  $Q$  表方法相比在时空性能上有了一定程度的提高。首先, 从时间性能上来看, 由于神经网络具有泛化特性, 只需学习部分状态 - 动作的  $Q$  值, 就

可以获得其它近似状态 - 动作的  $Q$  值, 因此其学习速度要比  $Q$  表方法快。其次, 从空间性能上来看, 神经网络所需的存储空间只和网络的连接权个数有关, 而连接权的个数一般要远远小于状态的个数, 所以采用神经网络来存储  $Q$  值函数的方法所占用的存储也要小于后者。

#### 5 结语

提出将神经网络应用于  $Q$  学习, 有效地提高了系统泛化能力, 实现了 RoboCup 中底层传球策略的优化。实验结果说明了与神经网络结合的  $Q$  学习在一定的训练周期结束后收敛, 最终得到了优化的传球行为序列。今后有关将 RBP 模型应用与上层踢球策略以及改进神经网络的映射能力加强  $Q$  学习有待进一步研究。

#### 参考文献:

- [1] Stone P. Layered learning in Multi-Agent System [D]. Pittsburgh, PA: Computer Science Department, Carnegie Mellon University, 1998.
- [2] Kaelbling L P, Littman M L, Moore A W. Reinforcement learning: A survey[J]. Journal of Artificial Intelligence, 1996, 4: 237 - 285.
- [3] Sutton R S, Barto A G. Reinforcement Learning[M]. Cambridge, MA: The MIT Press, 1998.
- [4] Tsitsiklis, John N. Asynchronous stochastic approximation and  $Q$  - learning [J]. Machine Learning, 1994, 16(3): 185 - 202.
- [5] 叶世伟, 史忠植. 神经网络原理[M]. 北京: 机械工业出版社, 2004.
- [6] 丛爽. 面向 MATLAB 工具箱的神经网络理论与应用[M]. 合肥: 中国科学技术大学出版社, 2003.

(上接第 62 页)

#### 5 结语

论文的创新之处在于提出了基于 Internet 的教育资源网格体系结构, 采用 RDF 描述教育资源, 并提出索引表和链表结合的方式组织教育资源。链表和索引表结合的资源组织方式, 是根据《基础教育教学资源元数据规范》提出来的, 符合教育资源的特征。同时, 它既适应教育资源网格环境的动态性、异构性, 又能减少查询的成本, 是较好的教育资源组织方式。索引表和链表的动态更新策略, 将是以后研究的重点。

#### 参考文献:

- [1] 朱莹, 吴军华, 汪婷婷, 等. 网格资源描述技术的比较研

究[J]. 微计算机信息, 2006(22): 176 - 178.

- [2] Li Juan, Vuong S. A semantics - based routing scheme for grid resource discovery[C]// e - Science and Grid Computing, 2005. First International Conference. [s. l.]: [s. n.], 2005: 438 - 445.
- [3] CELTS - 42 CD1.6. CELTS - 42 - 2002. 教育信息化技术标准[S]. 2002.
- [4] Cai M, Frank M R, Chen J, et al. MAAN: A multi - attribute addressable network for grid information services[C]// Grid Computing, 2003. Proceedings. Fourth International Workshop. [s. l.]: [s. n.], 2003: 184 - 191.
- [5] Tang C, Dwarkadas S. Hybrid Global - Local Indexing for Efficient Peer - to - Peer Information Retrieval[C]// Proceedings of USENIX NSDI. [s. l.]: [s. n.], 2004.