

语义网格的基础理论、模型与方法研究进展

◆ 诸葛海

摘要 国家重点基础研究发展计划(973计划)项目“语义网格的基础理论、模型与方法研究”针对互联网信息有效共享的基础科学问题进行了研究,旨在突破现有数据模型的局限,建立新的网络资源有效共享的模型、理论和方法。本文主要介绍了该项目启动3年来在建立语义资源空间模型的理论和方法方面取得的一些重要进展。

关键词: 语义网格 资源空间模型 语义链 网络 知识流 对等网

一、语义网格

作为一种全球范围的信息共享设施,互联网已成为推动经济和社会发展的动力。但目前的互联网应用平台还不能满足人类社会对网络资源(信息、知识和服务)有效共享和管理的需求。一个重要原因是当前的互联环境缺乏机器可理解的语义信息,难以提供有效的智能服务。为此,人们正在试图通过提出互联网语言和建立本体等方法来建立语义互联网^[1]。

资源的有效管理是计算机和网络有效工作的基础。计算资源管理发展的里程碑包括文件系统、数据库和操作系统。关系数据模型是一个重要里程碑,它实现了数据的规范化和有效管理。截至目前,数据库领域已产生了3个图灵奖。但经典的数据库模型是在集中式背景下提出来的,当时主要是解决对简单和规范数据管理的问题,并不考虑对复杂、异构、海量、分散、动态的网络资源的有效共享和管理。

实现有效信息、知识和服务共享和管理的一条重要途径是建立适用于互联网环境的语义数据模型,能够对内容进行有效的管理。语义网格是面向语义的智能互联应用环境。它需要建立一套新的理论和方法来实现对地理分布的各种网络资源进行规

范和灵活的组织,实现网络资源的有效共享和智能利用^[2-4]。

二、核心科学问题

本项目从互联网的自然特性、人类社会对互联网应用的需求、信息科学和互联网本身的进化需求所产生的科学问题出发,研究新一代互联网应用平台的关键理论、模型和技术。项目重点研究和解决以下3个关键科学问题。

1 兼备规范组织和自组织的资源组织模型

规范组织是管理资源的有效方法,但互联网应用要求自组织性。因此,如何协调规范性和自组织性是互联网时代资源组织的关键问题。本项目通过定义一套语义空间的范式来规范重构各种网络资源并与对等网络技术相结合,使动态网络环境下的无序资源规范化,使资源操作准确、方便,以实现有效的资源共享^[5-6]。

2 语义互联问题

如何实现分布在全球的各种网络资源在语义层上互联,消除资源语义孤岛。通过简洁的方式从资源空间映射到语义空间,可实现异构资源在语义层的互联。基于此,本项目提出了语义链网络模型来实现自主的语义互联,通过与资源空间模型相结合,形成单一语义映像,使得各种资源在简洁的语义空间中得到统一和互联^[4,7-8]。

3 智能聚融问题

如何解决互联环境这一复杂系统的自组织和优化从而使系统变得更有效力。智能耦合可使各种资源(信息、知识和服务)动态耦合起来为某种应用提供智能服务。但紧耦合的优化并不符合互联社会所要求的自主性。为此,我们提出了知识流模型,以松耦合方式实现这种优化^[9-10]。

三、重要进展

1 资源空间模型研究进展

分类是人类认识世界和管理资源的最基本的方法,它也有助于我们管理网络资源。为此,我们系统地提出了基于分类语义的资源空间模型^[11],其创新

本文作者: 诸葛海 研究员, 中科院计算技术研究所

研究资助: 国家重点基础研究发展计划(2003CB317000)

点如下:

(1) 通过规范多维分类空间, 提出了资源空间模型的范式理论及其完整性约束, 为准确、正确的资源操作提供了理论保障;

(2) 为处理分类和资源操作的不确定性提出了概率资源空间模型;

(3) 有机结合对等网的可扩展性和资源空间的规范性, 提出了基于对等网的分散式资源空间模型;

(4) 为了支持有效的资源查询, 提出了适用于资源空间的物理存储机制;

(5) 为了深刻理解资源空间模型的内在特性, 提出了资源空间搜索的复杂性理论, 资源空间模型查询操作的代数和演算理论, 以及资源空间模型查询操作的完备性理论;

(6) 提出了资源空间模型和语义链网络的集成及其关系理论以及资源空间模型、互联网本体语言(OWL)和关系数据库间的转换和集成方法。

资源空间模型可望成为互联网资源管理的重要基础设施。

2 互联语义计算模型研究进展

互联网的成功关键在于其简洁的超链结构设计和使用模式。为在继承的基础上发展互联网, 我们提出了基于关系语义的语义链网络模型以及语义链的自动发现机制, 建立了资源空间模型到语义链网络的映射、语义链的代数模型、语义关系自动发现的方法, 以及结合类比、归纳和演绎的语义关系推理方法^[2-4,7], 从而构建了独特的语义互联网的模型、理论和方法。目前, 资源空间模型和语义链网络模型已在敦煌文化展示领域得到了应用, 并获得了多项知识产权。

3 语义覆盖网和知识流研究进展

通过在大规模动态网络上构建语义覆盖层来支持智能应用是构建未来智能互联环境的一个重要问题。为此本项目提出了可扩展的结构化对等语义覆盖网模型和构建方法以及基于结构化和非结构化两种对等网模型的分布式资源空间模型和方法。为实现自主的知识共享, 提出了对等知识流模型, 揭示了知识共享的动态本质和社会性。同时也有有机结合知识流网络、语义链网络和对等网络, 构成了语义网格的高层体系, 为解决大规模网络环境中资源的智能聚融和自组织提供了新方法^[2,3,6-8,11-16]。

相关研究成果已发表在本领域及相关领域著名期刊 IEEE Computer Communications of the ACM, IEEE TKDE, IEEE TPDS, IEEE Transactions on PAMI 和 ACM Transactions on Internet Technology 等上, 以及

WWW, VLDB, HIPS 等顶级国际会议上, 并应邀在系列国际会议做特邀报告 40 余次。同时, 出版了 The Web Resource Space Model 和 The Knowledge Grid 两部英文专著。此外, 项目发起了系列国际会议 SKG(目前已成功举办了三届), 还主办了 14 次国际会议。在 ACM Transactions of Internet Technology 和 IEEE TKDE 等国际著名刊物上组织了 10 次专刊。

项目的研究成果已在国际上产生重要影响, 并推动了相关领域的发展。项目中关于智能聚融的研究在 Communications of the ACM 发表后得到 Science 的评论; Web 发明人 Tim Berners-Lee 在其重要论著 A Framework for Web Science 中引用了本项目关于语义链网络的论文; 项目研究成果还得到国际刊物 Concurrency and Computation: Practice and Experience 主编 Geoffrey Fox, Web Semantics 主编 Carole Goble, 英国皇家计算机学会主席 Nigel Shadbolt, 语义网格倡导者之一 David De Roure, 对等计算专家 Karl Aberer, IEEE 会士 Toru Ishida 和 D. S. Yeung 等专家的引用和正面评价。SC 检索表明: 项目发表的论文数和被引数在知识网格方向均排名第一。据 Thomson Essential Science Indicators, 本项目的论文引用率位居领域 Top 1%。

四、下一步工作

下一阶段研究的重点是结合互联网的发展, 开展资源空间模型系统的研发, 完善资源空间模型和语义链网络模型的理论体系, 提出资源空间模型 2.0 版, 强化语义网格技术在文化和科学等领域的应用研究。语义网格致力于建立一种基于语义的智能互联应用环境, 有望在原创性的理论、关键技术以及应用方面产生突破, 形成新的研究领域, 产生重要的国际影响。

参考文献

- [1] Berners-Lee T, Hendler J, Lassila O. Semantic Web. Scientific American, 2001, 284 (5): 34-43.
- [2] Zhu G H, Shi X. Toward the eco-grid: a harmoniously evolved interconnection environment. Communications of the ACM, 2004, 47 (9): 79-83.
- [3] Zhu G H. Semantic Grid: Scientific issues, infrastructure and methodology. Communications of the ACM, 2005, 48 (4): 117-119.
- [4] Zhu G H. The future interconnection environment. IEEE Computer, 2005, 38 (4): 27-33.
- [5] Zhu G H. The Knowledge Grid. World Scientific Publishing Co., Singapore, 2004.

- [6] ZhuGe H, Sun X, Liu J, Yao E, Chen X. A scalable P2P Platform for the knowledge grid. *IEEE Transactions on Knowledge and Data Engineering*. 2005. 17(12): 1721—1736
- [7] ZhuGe H, Liu J, Feng L, Sun X, Pei C. Query routing in a peer-to-peer semantic link network. *Computational Intelligence*. 2005. 21(2): 197—216
- [8] ZhuGe H. Autonomous semantic link networking model for the Knowledge Grid. *Concurrency and Computation: Practice and Experience*. 2007. 7(19): 1065—1085
- [9] Liu J, Jin X, Wang Y. Agent-based load balancing on homogeneous mingrids: macroscopic modeling and characterization. *IEEE Transactions on Parallel and Distributed Systems*. 2005. 16(7): 586—598
- [10] ZhuGe H. Discovery of knowledge flow in science. *Communications of the ACM*. 2006. 49(5): 101—107
- [11] ZhuGe H. *The Web Resource Space Model*. Springer. 2007
- [12] ZhuGe H. China's e-science knowledge grid environment. *IEEE Intelligent Systems*. 2004. 19(1): 13—17
- [13] ZhuGe H. Exploring an epidemic in an e-science environment. *Communications of the ACM*. 2005. 48(9): 109—114
- [14] ZhuGe H, Li X. Peer-to-peer in metric space and semantic space. *IEEE Transactions on Knowledge and Data Engineering*. 2007. 19(6): 759—771
- [15] ZhuGe H, Chen X, Sun X, Yao E. HRing: a structured P2P overlay based on harmonic series. *IEEE Transactions on Parallel and Distributed Systems*. June 2007. <http://doi.ieeecomputersociety.org/10.1109/TPDS.2007.70725>
- [16] ZhuGe H, Feng L. Distributed suffix tree overlay for peer-to-peer search. *IEEE Transactions on Knowledge and Data Engineering*. <http://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=69>
- [17] ZhuGe H, Guo W, Li X. The potential energy of knowledge flow. *Concurrency and Computation: Practice and Experience*. 2007. 19

2067—2090

- [18] ZhuGe H, Sun X. A virtual ring method for building small-world structured P2P overlays. *IEEE Transactions on Knowledge and Data Engineering*. accepted

Research Progress on the Fundamental Theory Model and Method on the Semantic Grid

ZhuGe Hai

Institute of Computing Technology

Chinese Academy of Sciences, Beijing 100080

The Semantic Grid aims at three fundamental scientific issues of future interconnection environment: normal organization, semantic interconnection and intelligent clustering. This paper reports the latest research progress of the National Basic Research Program's Semantic Grid Project. It intends to break through the limitation of typical data models to form a new semantic model and relevant theories for efficiently sharing network resources. It is to establish a semantic-rich intelligent interconnection environment. Important progress has been made by three years' effort, especially in research on the Web Resource Space Model and the Semantic Link Network model.

Keywords: semantic grid; resource space model; semantic link network; knowledge flow; peer-to-peer network

科学人

65位科学家获 2007年度何梁何利基金奖

2007年度何梁何利基金颁奖大会 2007年 10月 31日在京举行。桥梁工程专家林元培和油田勘探专家马永生两位专家获何梁何利基金科学与技术成就奖;乐嘉陵、彭实戈、陈和生、郑厚植、陈难先、侯建国、杨玉良、支志明、魏奉思、杨经绥、黄润秋、方精云、孟安明、罗辽复、武维华、吴明珠(女)、陈焕春、祖元刚、蒋华良、胡之璧(女)、曾益新、尚永丰、强巴赤列、郭亚军、韩德民、陈孝平、陈均远、叶声华、方滨兴、王 圩、吴佑寿、怀进鹏、孙优贤、都有为、俞耀庭、徐匡迪、陈志杰、张恩和、贺祖明、赵宪庚、张晓锋、张永维等 42名科学家获何梁何利基金科学与技术进步奖;陈剑平、蔡秀军、凌沛学、刘昌胜、马 骋、刘 峰、李殿中、邢子文、高金森、郭三堆、刘积仁、孙朝晖、孙克勤、李世鹤、孙逢春、叶志翔、钱 锋、徐一戎、哈木拉提·吾甫尔、李 健(女)、孙龙德等 21名科学家获何梁何利基金科学和技术创新奖。