

语义网格:语义 Web 与网格计算的融合^{*})

吴 磊 都志辉

(清华大学计算机科学与技术系 北京 100084)

摘 要 本文对近几年 Web 服务、语义 Web 及网格计算等一些新兴的技术进行了简要的回顾,并对它们的背景、特点及相互关系进行了分析,在这基础上,介绍了一种新的网格发展趋势——语义网格,对它的背景、目标、体系结构及其知识层进行了详细描述并简要总结了当前语义网格的研究现状。

关键词 Web 服务,语义 Web,网格计算,e-Science,语义网格

Semantic Grid:the Mergence of the Semantic Web and Grid Computing

WU Lei DU Zhi-Hui

(Department of Computer Science & Technology, Tsinghua University, Beijing 100084)

Abstract The field of Grid Computing has now evolved to a broader definition, this is closely related to the Semantic Web vision. In this paper, firstly, we review some new technologies emerging recent years, such as Web Service, Semantic Web and Grid Computing, discuss the relationships among them, then we explore a new concept——Semantic Grid with the detail description of it's background, definition, architecture and knowledge layer. Finally, we draw a conclusion to the Semantic Grid.

Keywords Web service, Semantic web, Grid computing, e-Science, Semantic grid

1 引言

信息技术的发展日新月异,互联网作为人们沟通信息和协同工作的有效工具,正深刻地影响并改变着人们的生活。传统的互联网实现了计算机硬件的连通,人们可以利用它进行数据传输,例如 E-Mail, Ftp, 而 Http/Web 的出现实现了网页的连通,通过超链接,所有网上的脚本/文档都可以在同一个浏览器里访问。20 世纪 90 年代中期, Ian Foster 和 Carl Kesselman 提出了网格^[1]的概念,想要利用互联网把分散在不同地理位置的电脑组织成一台“虚拟的超级计算机”,实现所有资源的连通和共享。

虽然网格被认为是通往下一代互联网技术的途径,但是目前,就技术层面而言,网格计算只是众多的网络计算模式中的一种,其他当前的热点技术包括:面向对象的分布式计算技术(COM/DCOM, CORBA 等)、语义 Web、Web 服务、对等计算、普及计算等,本文第 2 节将对这些技术产生的背景进行剖析,并对它们的技术特点和相互关系做出分析。

2001 年,英国工程与物理科学委员会和工贸部的核心 e-Science(“e-Science”,通常是一个应用于支持科学家的高级计算技术的术语。e-Science 为计算机和通信技术如何知识并增强科学进程提供了一幅希望蓝图,它允许科学家以一种更有效的方式生成、分析、共享和讨论他们的观点、经验以及结果,并能提供高度易于使用的无缝自动技术,还有全球范围内灵活的协作和计算^[2]。)计划提出了语义网格的概念,目标是对语义网和网格进行融合^[2]。语义网格被认为是下一代 e-Science 的基础设施,它试图将语义 Web 的思想和技术融入现今的网格,GGF 也成立了专门的语义网格研究小组来提供对语

义网格^[3]的支持。本文第 3 节将介绍语义网格的背景、定义及研究现状,对语义网格面向服务的体系结构进行描述,并对其其中的一些关键技术进行分析。最后做出总结。

2 当前技术回顾

2.1 当前技术发展背景

当前的互联网上存在着极其丰富的计算资源、数据资源、软件资源等,其潜在的数据存储能力、计算能力非常巨大。科研学术界迫切需要一种更有效的资源共享及跨地域的协同工作、解决复杂问题的方式,而电子商务的急剧升温则促使工商企业界寻找一种途径对企业内及不同的企业间的商务系统进行整合,以最小的代价适应新的改变。因此,正是这种巨大的需求驱使人们不断思考和尝试,新的技术也就在这种背景下应运而生。

首先,可扩展标记语言 XML(eXtensible Markup Language)的诞生克服了 HTML 只适用于人机交互的弱点。作为一种技术或者说思想,XML 充分体现了标记语言的精髓,它对整个 IT 行业带来的影响是革命性的,它深刻地改变了人们描述、获取和使用信息的全过程。

同时,人们针对面向对象的分布式计算的弱点,提出了面向服务的体系结构 SOA,它能够提供平台来构建具有以下三个特征的应用程序服务:松散耦合、位置透明、协议独立。一个 SOA 系统由一组服务(Service)组成,每个服务都是可以向外提供明确定义的服务的自治系统。

此外,当前“我们拥有存储和访问数据的技术,但是我们却缺乏足够的将‘数据坟墓’(data tomb)(目前,随着因特网的蓬勃发展,因特网上的数据也正在以一种前所未有的

^{*} 本项目受北京市自然科学基金(4042018)、清华大学基础研究基金以及 IBM 基础研究基金的资助。吴 磊 硕士研究生,主要研究方向为语义网格。都志辉 副教授,主要研究方向为并行计算、网格计算。

速度急剧增加。在很多领域,数据消费的速度远小于数据产生的速度,很多数据被存储起来却从未被访问过,这就是所谓的“数据坟墓”)转化为有用的数据并从中提取出知识”^[4],这种需求意味着互联网将逐渐由信息和通信基础设施演变为知识传递基础设施,这对网格概念的内涵也产生了深远的影响。1999年,英国 CLRC (Central Laboratory of the Research Councils) 的 Keith G. Jeffery 首次提出了网格的三层概念抽象^[5],即计算网格/数据层、信息层、知识层,从下层至上层,对于数据的抽象层次越来越高。上层则建立在下层功能之上,需要控制和调用下层的服

综上所述,XML、面向服务的软件体系结构(SOA)以及网格三层概念模型的出现,为下一代互联网技术铺下了重要的基石。

2.2 Web 服务、语义 Web 与网格计算

回顾是为了更好地总结过去、面向未来。掌握了上述技术发展的背景,我们再来看看当前网络的几大热点技术:Web 服务、语义 Web 和网格技术。

2.2.1 Web 服务

Web 服务是继 COM/DCOM 和 CORBA 之后,一种新的分布式计算技术,是建立可互操作的分布式应用程序的新平台。

从技术观点上,Web 服务其实是由一群标准协议堆叠而起的、面向服务的架构。其中,XML 是 Web 服务平台中表示数据的基本格式,XSD 则是 Web 服务的数据类型系统;简单对象访问协议 SOAP 提供了标准的调用 Web 服务的 RPC 方法;Web 服务描述语言 WSDL 则对 Web 服务及其函数、参数、返回值进行描述,这是一种基于 XML 的语言;整个 Web 服务的底层采用 HTTP 通信协议,所以它能提供跨越防火墙的通信。

Web 服务之所以兴起并获得成功,是因为电子商务的急剧发展以及 B2B 计算领域的不断演进,Web 服务这种先进的面向服务的分布式计算技术能够让企业内部的应用软件整合以及企业之间的商务系统整合的成本大大降低,给企业带来了巨大的好处,给服务背后的 IT 系统提供了更大的灵活性。

2.2.2 语义 Web

WWW 之所以获得成功,是因为它以一种简便的方式(HTML、Http、URL、浏览器)将整个因特网上的文本、图像、声音等各种信息链接到一起,使用户能够很方便地搜索和浏览信息。如果 WWW 成为一个巨大的数据库,将人类易读的文档和机器可理解的数据以一种对人类和机器都很有效的方式连接起来,将会怎么样?

这就是 WWW 之父、W3C(World Wide Web Consortium) 总裁 Tim Berners-Lee 所设想的 Web 的将来。Tim Berners-Lee 及 W3C 的其他领导人最初在 2001 年 5 月提出了这个概念,随即该观点迅速成为 W3C 工作的重点,他们将它称之为语义 Web(the Semantic Web)。

Berners-Lee 对语义 Web 定义如下,“语义 Web 是对当前 Web 的扩展,其中信息被赋予了定义好的含义,能更好地支持计算机和人类的协同工作”。通过在现有的 Web 上增加元数据,语义 Web 将允许人类和机器以一种现在几乎不可能的方式发现和利用数据。

根据 Berners-Lee 的展望,语义 Web 将是一个能够理解人类语言的智能网络,可识别信息的意义,并对信息自动进行解释、交换和处理。它允许 Web 实体(软件代理、用户和程序)进行互操作,动态地发现和利用资源,提取知识并解决复杂的问题。

元数据对 Web 内容进行了注释和描述,我们可以利用它来实现计算机的操作,我们还需要用复杂、自动的操作给每一个 Web 资源赋予语义含义。语义网的分层模型包括^[4]:

- 一个 Web 资源集,拥有唯一的全球标识,由元数据以一种通用、共享的方式描述,并拥有一系列通过本体论(ontology)推断新的元数据和知识的规则。

- 一个基础服务集,例如基于元数据、本体论和语义搜索引擎的推理和查询,与现有因特网服务如 DNS、基于关键字的搜索相比,这些服务将会是极大的改进。

- 使用基础服务集开发的高层应用程序集。

在当前阶段,重点的研究方向是用于元数据和本体论标准建模和实现的语言和技术的开发,例如 XML Schema 和 RDF Schema, DAML+OIL 和 OWL(www.daml.org, www.w3.org)。

2.2.3 网格计算

通常,网格被认为是通往下一代因特网技术的途径。至今为止,网格的发展大致经历了三个阶段^[6]:

第一个阶段是 20 世纪 90 年代早期到中期,是网格的萌芽阶段,其代表性的两个项目是 FAFNER (Factoring via Network-Enabled Recursion) 和 I-WAY (Information Wide Area Year)。

网格发展的第二个阶段是以 Globus (GT1 与 GT2) 等为代表的许多来自学术界的研究项目,持续时间大概从 90 年代中期到 2000 年左右。网格在这一阶段的研究重点是解决通用的网格平台所面临的异构性、扩展性以及适应性等问题。从体系结构的观点上看,这一阶段的网格在异构的分布式系统、资源、用户及本地策略之间使用网格中间件技术作为“粘合剂”^[4],以此来向用户和应用程序提供同构和无缝的环境。在这一时期网格主要是针对大规模数据与计算问题的解决展开研究的,研究问题的重点主要还是在科研领域。

网格发展的第三个阶段大概从 2000 年左右到现在,是网格技术标准化的阶段。在这个阶段,网格的重心开始转向分布式全球协作、面向服务的方法以及信息层处理。体系结构也从五层沙漏模型演变成 OGSA (Open Grid Services Architecture),面向服务的模型是网格标准的核心内容,而对元数据(metadata)(元数据是关于数据的数据,描述了一个特殊的数据集何时、由谁怎样进行群集,以及数据的格式如何)的关注则是网格管理自动化思想的体现,网格的应用范围已经全面突破了科学研究领域的局限,进入了社会生活的方方面面,这也是网格目前最新的发展阶段。

2004 年,IBM、Globus 联盟和 HP 共同提出了 WSRF。WSRF 是在保持 OGSI 概念和功能的前提下,对 OGSI 的重写(refactoring)和发展,作为一种新的底层技术实现基础设施,它将代替 OGSI 为 OGSA 提供基于 Web 服务的广泛而强大的支持。WSRF 实现了网格与 Web 服务的融合,它既可以充分利用已有 Web 服务领域的各种成果,又吸纳了网格技术,可以支持网格的需求,为网格和 Web 领域的发展建立了一个共同的基础。

2.2.4 Web 服务、语义 Web 和网格计算的关系

先来看网格的演化史,从技术层面上分析,其体系结构经历了从五层沙漏结构到 OGSA 的转变,OGSI 也改进为 WSRF,从应用层面分析,网格的重心已逐渐由科研学术界向工商业界转移,这里面并不单纯是技术先进性在起作用,网格的这种演变史与商业需求、Web 技术的演变(XML、RDF (RDF (Resource

Description Framework)是语义 Web 众多的基础规范之一,建立在 XML 的上层,是描述元数据的通用框架,它为通过 Web 交换计算机可理解的信息的应用程序提供了互操作性)等)和工业界的介入(IBM 等)息息相关,也是一个技术不断发展与融合的过程,OGSA 采用了面向服务的架构,试图融合 Web 服务,但从 OGSI 的实践看来,并没有完全获得成功,于是在新出的 WSRF 技术中,彻底保留了原有的 Web 服务技术并完全将之融合。

下面来分析语义 Web 和网格的关系^[5]。通常,网格被认为是下一代互联网技术,而语义 Web 被认为是下一代 Web。网格和语义 Web 有很多相似之处,但是各自强调的重点有所不同:传统意义上,网格侧重于计算,但是语义 Web 更倾向于推理和证明。图 1 是 Norman Paton 画的一幅示意图,揭示了这两者技术之间的关系。传统的网格基础设施能扩展 Web 的计算能力,同时语义 Web 能帮助网格扩展更丰富的语义,因此,我们可以看图中虚线以上部分,网格将向着“更丰富的语义”的轴的方向进一步演化。事实上,作为语义 Web 的核心——元数据(metadata)在网格演化的过程中(网格服务出现之前)就已经在一些网格项目的体系结构中扮演关键角色。

很自然地,人们开始思考,是不是能将现有的网格和语义 Web 技术进行融合,事实上,这方面的研究已经展开了,详见本文第 3 节。

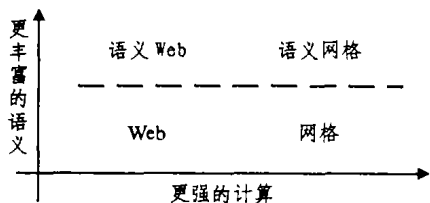


图 1 语义 Web 和网格

3 语义网格^[3]

3.1 语义网格的目标

2001 年,英国工程与物理科学委员会和工贸部的核心 e-Science 计划提出了语义网格的概念,目标是对语义网和网格进行融合^[2],GGF 也成立了专门的语义网格研究小组来提供对语义网格的支持。

语义网格被描述为一个开放的系统,其中,用户、软件组件和计算资源(由不同的组织拥有)动态构成其基础。它高度自动地在全球范围内支持灵活的协作和计算,而且,该环境可以针对不同的参与者进行个性化设置,还能在软件组件和其他相关的用户之间提供无缝的协作。

下面,我们将从两个方面介绍语义网格,一是其面向服务的体系结构,二是其对网格内涵在信息和知识层面上的延伸。

3.2 面向服务的体系结构

语义网格是一种面向服务的体系结构^[7],其中,实体(entity)可以通过不同形式的契约(contract)来相互提供服务(service)。首先,来明确几个概念:

服务(service):可以简单地被视为对一些内容或处理能力的特征抽象和封装;

所有者(owner):指一个负责提供服务供他人消费的实体(个体或机构);

消费者(consumer):指那些决定尝试或调用服务的实体;

服务契约(service contract):指定了所有者同意提供服

务给消费者的术语和条件。

所有的服务都有一个所有者(或所有者集),所有者为服务的存取设定术语和条件。所有者和消费者之间的联系是通过服务契约明确的。服务所有者之间在市场(marketplace)中进行交互,而建立并运行市场的被称为市场所有者(market owner)。

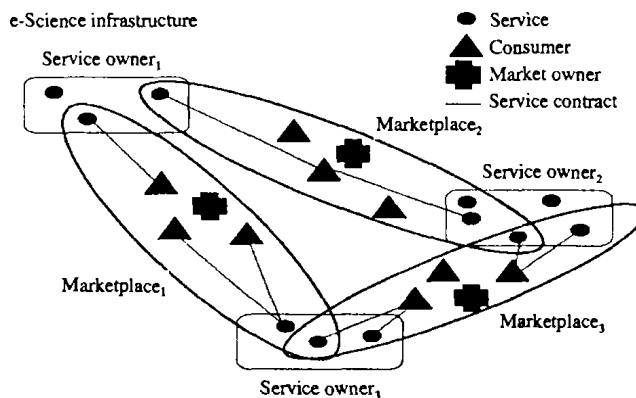


图 2 面向服务的体系结构:关键组件

总结一下,面向服务的体系结构的关键组件如图 2 所示:服务所有者(圆角矩形)在特定协议下(生产者和消费者之间的实线)提供服务(填充的圆形)给服务消费者(填充三角形)。每个所有者—消费者之间的交互发生在给定的市场(椭圆代表),规则由市场所有者(填充的十字)建立。市场所有者可能是市场中的一个实体(或者生产者或者消费者),或者是一个中立的第三方。

3.3 语义网格知识层

这是网格向语义网格演化的另一条主线。

和快速膨胀的 Web 页相比,人们始终缺乏有效的方式通过互联网来发布、组织、共享知识资源,在知识层面上,人们是相互孤立的,只能使用各自的知识来完成任务。知识管理在促进组织革新和提高生产力方面扮演着重要的角色。一个基于互联网的知识管理应该支持用户能随时存储他生成的有用的知识,并能方便地从分布在互联网上的知识仓库中获取需要的知识来解决遇到的问题。这样,随着因特网用户的增多,互联网上的知识资源能够迅速积累并发展为整个互联网团体可以通用的知识财产。在语义网格中,知识层的目标就是作为一个支持科学知识的管理和应用程序的基础设施,达到特定的目的和目标。它建立在由数据/计算层和信息层提供的服务上。

3.3.1 知识生命周期 可以被视为一组挑战集,也可以被看作一系列的阶段。每个阶段都可以从不同的侧面被视为瓶颈。知识的生命周期包括如下部分^[7]:

- 知识获取:抓取周围的信息,并将它转换为知识并使它可用。
- 知识建模:是搭建在知识获取和知识使用之间的桥梁。
- 知识检索:一旦知识被获取和建模,那么就需要存储在某些地方,从而可以有效地检索。
- 知识重用:知识构件必须经常进行重新构建,从而高效地使用知识。
- 知识发布(publish)或分发(disseminate):将知识传递到随后需要它们的人手中。
- 知识维护:保持现在的知识内容。

(下转第 30 页)

时,并不马上启动路由发现过程,而是把业务都转到剩下的路径上进行,而且两个路径间越互不相交,相互干扰及某段链路断开对两条路径带来的影响就越小。通过这种方式,限制了路由请求包的数量,同时降低了网络中冲突和与数据包竞争的次数,导致网络具有较好的吞吐性能和鲁棒性。从图3中可以看出,本方法较LAR协议在不同的暂停时间上包投递率都有一定的提高,尤其是在网络节点高速移动的情况下更明显。这是因为网络拓扑变化越频繁,链路越容易断开,本文提出的方法作用就越明显。

图4给出了在相同的移动情况下源端发出的路由请求包的数量。由于本文方法一次路由请求可能会得到两条不同路径,并且只在两个路径全部失效后才重新启动路由发现过程,因此源端总的路由请求的次数比单路径LAR少。而图5则表明,在路由控制开销指标的比较上,本文方法大大优于LAR。这是因为,一方面源端发出的路由请求包的数量降低了,大大减少了网络内各中间节点的转发数量。另一方面,尽管改变了转发规则,局部提高了转发包的数量,但由于采用了在请求包内携带转发节点地址信息的方式,提高了源端节点获知目的节点地址的概率,请求区域缩小,参与转发的节点数目降低了。同时,请求区域变为圆形区域,使路由由发现成功的次数增加了,也大大降低了二次广播的次数。

结论 本文提出了一种基于位置信息的Ad-Hoc网络(即MANET)多路径路由方法。在LAR路由协议的基础上通过改变路由请求包的转发规则,并在请求包内携带每个中间转发节点的位置信息,使目的节点采用BFS算法利用获得的中间节点位置信息来计算出另外一条不相交路径,通过把业务分配到两个完全不相交路径中,来实现负载均衡和增强网

络的鲁棒性。但本文只是把业务平均分配到两个路径中对方法进行验证,并没有充分讨论两个路径的QoS及安全情况,如何设计较好的业务分配算法以更好提高Ad-Hoc网络的性能将有待进一步研究。

参考文献

- Perkins C E, Royer E M, Das S R. Ad Hoc On Demand Distance Vector (AODV) Routing. IETF Internet Draft, draft-ietf-manet-aodv-13.txt, Feb. 2003
- Broch J, Johnson D, Maltz D. The Dynamic Source Routing Protocol for Mobile Ad hoc Networks. <http://www.ietf.org/internet-drafts/draft-ietf-manet-dsr-04.txt>, IETF Internet Draft, Nov. 2000
- Ko Y, Vaidya N H. Location-aided routing (LAR) in mobile ad hoc networks. In: Proc. of the Fourth Annual ACM Intl. Conf. on Mobile Computing and Networking (MobiCom 1998), 1998. 66~75
- Nasipuri, Das S R. On-Demand Multipath Routing for Mobile Ad Hoc Networks. In: proc. of the 8th Int. Conf. On Computer Communications and Networks(IC3N), Boston, Oct. 1999
- Wu K, Harms J. On-demand Multipath Routing for Mobile Ad Hoc Networks. In: Proc. of 4th European Personal Mobile Communication Conf. (EPMCC), Vienna, Austria, Feb. 2001
- Lee S J, Gerla M. Split multipath routing with maximally disjoint paths in ad hoc networks. In: Proc. of IEEE ICC'01, June 2001
- Wu K, Harms J. Performance Study of a Multipath Routing Method for Wireless Mobile Ad Hoc Networks. In: Proc. of IEEE/ACM 9th International Symposium on Modeling, Analysis and Simulation (MASCOTS 01), Cincinnati, Ohio, Aug. 2001. 99~107
- Sherlock R, Mooney P, Winstanley A, Husdal J. Shortest Path Computation: A Comparative Analysis. GISRUK, Sheffield, UK, April 2002. 91~94
- Bagrodia R, et al. PARSEC: A Parallel Simulation Environment for Complex Systems. IEEE Computer, 1998, 31(10):77~85
- UCLA Parallel Computing Laboratory and Wireless Adaptive Mobility Laboratory, GloMoSim: A Scalable Simulation Environment for Wireless and Wired Network Systems. <http://pcl.cs.ucla.edu/projects/domains/glomosisim.html>

(上接第3页)

3.3.2 本体库(Ontology)和知识层^[7] 本体库的概念对于取得表达能力是很必要的,在用知识进行建模和推理的过程中需要这种能力。总的来说,一个本体库决定了术语的外延和它们之间的关系。然而,在知识和Web工程的上下文空间中,一个本体库仅仅是一个已发布的或多或少取得一致的一些领域内容的概念。本体库可能描述对象、过程、资源、能力或者其它的东西。

本体库的好处是改进了在系统之间的通信,这些系统可以是机器、用户或者组织。它们的目标就是建立一个一致的,可能是规范的模型。它们努力保持一致,去除二义性,并集成了一系列的观点。接受本体库得到的另外一个好处是可互操作性,这也是为什么在语义网格中占有如此重要地位的原因。一个本体库可以作为一种国际语,它可以推动内容的重用,确保给出一个清晰的规范来表达内容或者服务是关于什么的,增加内容和服务被成功集成的机会。

可以认为,本体库提供了内容的通信、集成和共享的基础。但是它们还可以提供其他的好处。本体库可以通过去除模糊性和指出相关术语,或者通过建立从页面得到的信息与其他信息建立联系,来改进搜索准确性。还有,互联网推理系统开始出现,它通过从现有的Web中抽取和产生注释来开发本体库。

结束语 任何事物发展的背后,总有其深刻而本质的原因,拨开技术层面的面纱,我们可以发现,更丰富、更灵活、更方便的信息资源共享和大规模的协同工作才是真正的推动技术发展的本质所在。

科研学术界的美好设想、商业需求、工业界的策略使得技术的概念和内涵变得前所未有的模糊与难以界定,现今的诸

多技术如Web服务、语义Web、网格计算之间存在很多共性,但侧重点又各有不同,无论如何,技术总是发展、相互影响又最终走向融合,或者说找到一个在符合技术先进、满足各方利益等之间找到一个合适的平衡点。

目前,语义网格是个全新的概念,它试图实现语义Web和网格计算技术上的融合,成为未来e-Science的基础设施,一些围绕语义网格的项目也相继展开,如myGrid(www.mygrid.org.uk)、Comb-e-Chem(www.combechem.org)、Geodise(www.geodise.org)、CoAKTingG(www.aktors.org/coaktng)等e-Science项目。

语义网格所蕴含的这种思想有着极为广阔的发展空间,相信随着技术的深入研究和不断发展,语义网格将真正成为下一代智能网格,对人类的生活产生重大而深远的影响。

参考文献

- Foster I, Kesselman C, eds. The Grid: Blueprint for a New Computing Infrastructure. San Francisco, CA: Morgan Kaufmann Publishers, 1998
- De Roure D, Jennings N R, Shadbolt N R. Research Agenda for the Semantic Grid: A Future e-Science Infrastructure. [UkeS-2002-2, Technical Report of the National e-Science Centre]. 2001
- The Semantic Grid Community Portal. GGF. <http://www.semanticgrid.org/>
- Cannataro M, Talia D. Semantics and Knowledge Grids: Building the Next-Generation Grid. IEEE Computer Society, 2004
- De Roure D, Goble C. Semantic Web and Grid Computing. Sep. 2002. <http://www.semanticgrid.org/documents/>
- De Roure D, et al. The Evolution of the Grid. Grid Computing: Making the Global Infrastructure a Reality. In: F. Berman, A. J. G. Hey, and G. Fox, eds., John Wiley & Sons, 2003. 65~100
- De Roure D, Jennings N R, Shadbolt N. The Semantic Grid: A Future e-Science Infrastructure. Grid Computing: Making the Global Infrastructure a Reality. In: Berman F, Hey A J G, Fox G, eds. John Wiley & Sons, 2003. 437~470