



2007-2008 年国外科学工作流研究进展

Review of the Study on the Scientific Workflow Abroad from 2007 to 2008

苏明明 (中国科学院国家科学图书馆 中国科学院研究生院 北京 100190)

宋文 (中国科学院国家科学图书馆 北京 100091)

[摘要] 科学工作流为促进和保障虚拟知识环境下的科研协作提供了一种解决方案。2007 年以来,国外学者围绕科学工作流开展的研究主要涉及科学工作流管理系统开发和优化、来源管理、系统可视化、语言、数据流控制、建模等方面。在领域应用中解决具体问题及优化以往的研究成果是这一阶段科学工作流研究的突出特点。

[关键词] 科学工作流 系统 可视化 控制

[中图分类号] G250.7 [文献标识码] A

[Abstract] Under the Virtual Knowledge Environment, the scientific workflow provides a solution for the promotion of the research collaboration. Since 2007, studies on the scientific workflow abroad have focused the development and optimization of the scientific workflow management system, the management of the scientific workflow provenance, the visualization of the scientific workflow system, the scientific workflow language, the control of dataflow, as well as the modeling of the scientific workflow. At this stage, solving specific problems and optimizing the results of previous studies are the prominent characteristics of the scientific workflow research.

[Key words] Scientific workflow; System; Visualization; Control

数字化环境对科研人员的科研行为产生了巨大的影响,人们从事科学研究的方式和方法也在发生着改变,在虚拟环境下从事科研活动已逐渐成为科研人员工作的常态。科学研究的协作频率和效率在虚拟知识环境下得到提高,科学工作流正是其中一种促进和保障协作的技术路线,它提供了一个专门的程序环境以减少科研人员在计算科学环境中进行研究的投入^[1]。目前国内针对科学工作流的研究并不多,截至 2008 年末,CNKI 与维普合计收录的以“科学工作流”作为关键词或篇名的文章仅 8 篇,通过百度、谷歌等搜索引擎搜索到的中文相关资料也非常有限。相比之下,国外关于科学工作流的研究更多一些,本文拟对 2007-2008 年国外科学工作流的相关研究成果进行分析整理,力求反映该领域的研究现状和发展方向。

1 2007-2008 年国外科学工作流研究概述

1993 年国际工作流管理联盟 (Workflow Management Coalition, 简称 WfMC) 成立,在其推动下,目前业务工作流的研究已经趋于成熟。而相对来说,在业务工作流间接

影响下产生的科学工作流仍有很大的发展空间。科学工作流之所以不能完全借用业务工作流已经成熟的技术、标准,主要是因为:业务工作流是用以控制流导向的,关注的是事务处理过程中业务活动的变化;科学工作流是用以控制数据流导向的,关注的是科学数据和其他信息在科研流程中各节点的差异变化对科研的影响。

近几年来,随着 E-Science、E-Research 等虚拟科研环境逐渐融入到科研活动中,科学工作流的相关研究也逐渐增多。据不完全统计,2007 年以来涉及科学工作流研究的国际会议至少有 25 次。纵观这些国际会议的主题可以明确:当前关于科学工作流的研究是 E-Science 研究领域关注的重点内容之一,研究的重点是围绕科学计算、网格等展开的,主要由计算机科学领域的专家从事这方面的研究,而且研究的内容从整体上可以分为技术研究和应用研究两类。技术研究的内容主要包括通用型科学工作流管理系统的开发和优化以及科学工作流的建模、可视化研究、来源管理、语言研究与控制研究等几个方面;应用研究主要是

研究科学工作流在具体应用环境中的解决方案。需要强调的是,科学工作流的技术研究和应用研究经常是相互依赖出现的,即技术研究以实际应用的需要为出发点,应用研究通过技术方案的改进而实现。

2 2007-2008 年国外科学工作流研究的主要方面

2.1 新型科学工作流管理系统不断涌现

科学工作流管理系统是对多领域或专门领域科学工作流进行管理的系统,它的研发降低了科研人员进行领域科学工作流管理的成本。目前已经开发的科学工作流系统有 Taverna、Kepler、Triana、GridNexus 等,由于目前并不存在一个绝对优越的科学工作流管理系统,通常这些系统在实际应用到具体领域时都需要进行改造。

2.1.1 通用型科学工作流平台

通用型科学工作流平台是可复用于多学科领域的系统,其中以 Kepler 最具代表性。近两年 Barga R 等人新提出的 Trident 科学工作流平台^[2]也是典型的通用型科学工作流平台。Trident 是通过扩展商用领域的工作流系统的功能开发出来的,可以储存和利用本地资源和云计算中的分布式资源,并在多簇工作流中提供资源出处、监测、日志、时序安排等服务。Trident 的目标是:到 2009 年初将其作为开源平台推广应用于 E-Science 项目研究,并不断进行拓展,使其能够支持新的工作流模式和服务。Trident 目前主要应用于科学计算集群中,并与多个 E-Science 项目进行了合作,其中较为重要的是与 Pan-STARRS 项目的合作。

2.1.2 面向专门领域应用的科学工作流系统

与通用型科学工作流系统不同,面向专门领域的科学工作流系统在实际应用和所针对的工作流对象方面比较单一。IVIP (Intelligente Vernetzung verteilter Informationsquellen zur betriebs- und standortspezifischen Planung der Energiepflanzenerzeugung 用以支持能源作物基于地理位置的生产计划的业务规范信息源智力集成系统)和 SWARM (Scientific Workflow for Supporting Bayesian Approaches to Improve Metabolic Models,支持用贝叶斯概率方法提升代谢模型质量的科学工作流)是面向专门领域应用的科学工作流系统的两个典型代表。IVIP 是 Web 空间决策支持系统 (Web-based Spatial Decision Support System,简称 WSDSS) 的原型系统,它演示了使用土地分配和土壤质量等数字地理信息进行基于地理位置的决策制定的优越性^[3]。Tuot CJ 等人通过扩展 Kepler 成功地实现了 IVIP WSDSS 模型库,使农场主和区域专家能够依靠空间数据决策支持系统制定基于地理位置的农业种植计划^[3]。

代谢模型是用来分析基因序列的方法模型,SWARM^[4]正是支持用贝叶斯概率方法创建代谢模型的一种科学工作流环境。SWARM 具备分布式数据资源集成、数据格式转

换、数据更新、数据来源跟踪等功能,能将整个代谢模型的建模过程进行流线化。首先,从多资源集中抽取数据;然后,生成训练预测数据的训练集,并用贝叶斯概率技术聚合预测数据,随后对预测数据进行训练;最后,根据预测结果填充目标数据集的缺失数据,进而自动提升代谢网络的数据质量。通过贝叶斯概率推断的方式,SWARM 可以以较低的消耗在短时间内生成较多的基因代谢模型。

2.2 系统优化方案研究

当前大多科学工作流管理系统的使用都很复杂,而且在使用过程中常暴露出一些不适用的情况,因此对原有系统的优化改造是科学工作流研究的重点。

2.2.1 系统领域应用优化

近两年科学工作流应用的相关研究有一个显著的特点,即多数应用是结合应用领域的工作流特征通过对 Kepler 的领域化改造实现的。关于管理和分析海底电缆传感器数据问题就是 Kepler 改造应用的一个案例,该项研究是实时环境处理 (Real-time Environment for Analytical Processing,简称 REAP) 项目的一部分,研究中 Jones MB 等人根据数据涡轮环形缓冲网络总线缓冲数据矩阵,利用 Kepler 直接从该环形缓冲区访问数据,实现了对海洋数据进行有效、透明的实时事件探测,同时也将 Kepler 应用于海底电缆传感器工作流分析和数据存档等领域^[5]。对 Kepler 改造应用的另一个成功案例是 Podhorszki N 等人根据等离子体边缘仿真项目的需求,开发了应用于 Kepler 的组件,实现了针对等离子体融合仿真数据的工作流操作,如通过一个已建立的 Ssh (Secure shell,安全壳) 连接执行多个远程活动、通过 Ssh 进行远程文件处理、针对文件的工作流处理、使工作流重新启动的检查点处理、工作流组件的通用注册等^[6]。

2.2.2 多系统优化集成

多系统优化集成的研究包括两方面内容:一是集成不同系统的优点,二是实现不同系统间的互操作。在当前多科学工作流管理系统并存的情况下,将不同系统的优点进行集成并形成可在一定程度上进行互操作的科学工作流管理系统,是相关研究人员期望实现的问题。Mandal N 等人将 Kepler 的可视化构件与 Pegasus 的管理和映射构件进行集成,以方便科研人员充分利用两个系统的优点^[7]。但是该项研究只是在有限的方面实现了两个系统的集成,如果用户想在两种系统之间实现完全的互操作,在目前的技术条件下并不是一件容易的事。

2.2.3 无需异构格式转换的系统优化

科学工作流在很多情况下处理的是异构数据和资源,如果在工作流执行中强制对异构格式进行归并处理,将影响科学工作流管理系统执行的效率,甚至可能导致数据或资源的损坏,因此实现异构数据资源集成也是科学工作流管理系统的重要功能之一。Wei Tan 等人针对美国国家癌症学会

的生物医学信息网格项目的面向服务的基础设施 caGrid 进行分析, 根据 caGrid 的工作流需求, 在 Taverna 工作流系统的体系结构中对异构信息资源进行集成, 这样就无需将异构数据和软件进行格式转换即可对网格中的数据和软件进行无障碍操作, 实现了对异构数据资源的优化整合^[8]。

2.3 科学工作流来源管理研究

来源自动跟踪和来源信息储存是科学工作流系统的主要功能, 可靠的来源管理技术是生成可信赖的来源记录的重要保障, 这些记录保证了来源信息的真实性、可信赖性和有效性。

2.3.1 科学工作流来源安全管理研究

2007-2008 年关于科学工作流来源问题的研究主要集中在可靠的来源管理问题上。例如: Gadelha J 等人提出的 Kairos 方案是一种在网格环境下对著作权信息和时效信息进行保护的架构, 可用于知识产权声明、科学实验的时序排列揭示等方面^[9]。Cohen-Boulakia S 等人针对查询出处时出现的一些错误问题深入分析了如何从用户使用的角度设置来源信息颗粒度的方法, 提出了基于用户观点的来源寻址模型 ZOOM*User Views。该模型是科学工作流的一种简单而普通的来源模型, 可以完整表达数据来源信息, 为保证信息可靠性提供来源查询支持^[10]。Belhajjame K 等人提出了通过 User Views 技术关注用户关注的来源信息的方法用以管理嵌套的科学数据来源, 并将工作流元数据分为生成、使用和消亡等几个阶段, 分别对不同阶段信息进行捕捉, 以辨别不同来源的相似的数据产品^[11]。

2.3.2 来源元数据管理研究

为了便于分布式信息的安全交流, 近两年来科学工作流系统中的元数据管理主要研究的是来源元数据管理。Chebotko A 等人提出了一种用 RDBMS (Relational Database Management System, 关系数据库管理系统) 的存储和查询功能实现来源元数据管理的方法, 此种方法具有互操作性、可扩展性和语义推理能力。该项目研究主要做了 3 个方面的努力: 一是设计了两个 schema 映射算法来映射一个任意的 OWL (Web Ontology Language, 网络本体语言) 来源本体到关系型数据库 schema 之中, 扩展了一般的来源查询功能; 二是设计了 3 个有效的数据匹配算法, 根据生成的关系数据库 schema 映射来源 RDF (Resource Description Framework, 资源描述框架) 元数据到关系型数据中; 三是设计了一个与 schema 无关的 SPARQL-SQL 转换算法^[12]。

2.4 科学工作流系统可视化研究

可视化是当前计算机科学领域研究的一个热点问题, 近两年科学工作流的可视化研究重点关注可视化应用基础结构和可视化系统开发两个方面。

2.4.1 Sedna 基于 BPEL 的科学工作流可视化环境

Wassermann B 等人针对 BPEL (Business Process Execution Language, 业务过程执行语言) 面向业务工作流、缺少针对科学工作流应用的问题, 从简化科学工作流建模方法以便非领域专家也能进行工作流建模的角度开发了一种可视化语言和可视化建模环境 Sedna^[13]。这种建模环境在保证科学工作流继承 BPEL 规范的同时, 结合了其他语言和工具的优越性, 使科学家能够更便利地掌控科学工作流。

2.4.2 VIEW 可视化科学工作流管理系统

VIEW 是 Lin Cui 等人开发的一种科学工作流管理系统, 它集成了互操作、可扩展、语义推理、关系数据库管理系统的查询储存和可视化等功能或技术^[14]。它的特色主要表现在两个方面: 一是使用语义网技术揭示、储存和查询来源元数据信息; 二是以二维或三维图等方式揭示来源信息。VIEW 包括 6 个子系统: 工作台面系统、工作流引擎、任务管理器、数据产品管理器、来源管理器、工作流监测等。

2.5 科学工作流语言研究

目前已经开发出了多种工作流语言, 如应用范围比较广的 BPEL 业务过程执行语言是一种基于 XML 的、描写业务过程的编程语言, 它所描写的业务过程的每个步骤都由 Web 服务来实现^[15]; YAWL (Yet Another Workflow Language, YAWL 工作流语言) 是基于严格的工作流模型分析的工作流语言, 它支持所有的工作流模型并具有正式的基于 Petri 网的基础^[16]; XPDL (XML Process Definition Language, XML 过程定义语言) 是 WfMC 推荐的一种标准语言^[17], 等等。但是这些语言主要是面向业务工作流的, 由于科学工作流与业务工作流有明显区别, 所以需要对这些语言按照科学工作流的数据流特点进行改造或重新开发, 以便应用到科学工作流中。

2.5.1 突出并行设计的科学工作流语言

Martlet 是一种科学工作流语言, 也可作为一种程序设计模型, 它不仅允许编写并行程序, 而且可以在无需考虑并行细节的情况下分析分布式数据^[18]。Martlet 由函数编程得到启示, 通过构件将并行计算和数据分裂抽象出来。这种科学工作流语言允许在程序运行时自动调节并匹配数据集、在可获取资源的抽象描述环境下编写程序。Martlet 实现了在分布式数据集中编写程序、运行复杂计算以及以更加直观的方式在分布式系统中工作^[18]。

2.5.2 针对数据流的正式图形化工作流语言

比利时等国的科研人员证明了数据流是根据一系列规则在等级结构模式下构造出来的, 并且他们提出了 semisound 的概念, 即输入节点以单令牌环 (可能代表了一个复杂的数据集) 出发, 输出节点也以单令牌环 (代表了输出数据集) 结束, 在这种情况下, 不管如何计算都不会产生数据碎片^[19]。在这种基本原理的支持下, Gadelha J 等

人联合提出了DFL(DataFlow Language,数据流语言)。DFL是一种针对数据流的正式的图形化工作流语言,是对Petrinets和NRC(Nested Relational Calculus,嵌套关系演算)的扩展。DFL在令牌无嵌套历史管理、配置和状态转换等方面进行了改进,具有相对优越的转换系统语义的能力^[20]。

2.6 科学工作流控制研究

科学工作流控制是针对科学工作流运行过程中可能出现的异常进行处理的操作,包括科学工作流系统的异常控制、信息繁殖控制、时间控制等。

2.6.1 科学工作流系统的异常控制

Tavares T等人针对科学工作流系统的容错能力设计了一种使用了中间件Anthill和Mobius的框架^[21]。该框架主要由3个部件组成:可共享的知识库、程序管理器和运行时环境。该框架中引入了稳定的持久存储管理器PSM(Persistent Storage Manager)、工作流元数据管理器WMDM(Workflow Meta-Data Manager)和记忆数据储存库IMDS(In-Memory Data Storage),以提升工作流系统的容错能力。该框架在工作流运行过程中不断记录日志信息,并使用工作流元数据管理器对日志信息进行管理,通过基于并行虚拟机容错探测包的Anthill基础结构探测错误,一旦发现错误,系统将分别在产生错误的用户应用过滤器、工作流元数据管理器、储存库系统这3个点上数据进行恢复。

此外,科学工作流的分布式处理可能会导致来自于应用层、设置层、资源管理层等不同层面的异常,从目前的技术手段来看,探测这些异常继而动态地调整工作流仍然相当困难。针对这种情况,Tolosana-Calasanz R等人提出了两种异常处理模式,即异常繁殖模式和异常治理模式^[22]。

2.6.2 使用层次状态机解决信息流繁殖控制问题

可无限繁殖是信息的一个特点,科学工作流中的信息流也具备这样的特征,但是无限繁殖和错误繁殖会对科学工作流的正常运行产生负面影响。针对信息流控制问题,Yang Ping等人提出了使用层次状态机进行科学工作流建模,进而在科学工作流环境中对信息的繁殖进行控制的解决方案^[23]。科学工作流层次状态机(Hierarchical State Machine for Scientific Workflows,简称HSMSW)模型不同于传统的层次状态机模型,体现在:HSMSW使用连接渠道模仿模块;HSMSW转换动作并不严格按照某种格式进行;HSMSW扩展了层次状态机的变量作用域^[23]。

2.6.3 科学工作流的时间序列优化

在科学工作流系统中,为确保及时完成工作流,实例时间的一致性尤其重要的,设置时间限制可以相对有效地监测和保证时间的一致性。但是当前采用用户设定时间限制的系统中很少考虑系统的运行效果,因此可能导致频繁的时间错误,进而影响工作流的执行效果。Liu Xiao等人提出了一种对工作流活动期内不同时间节点上的权重进行

设置的方法,即设置时间粗粒度(coarse-grained)和细粒度(fine-grained)限制的策略,有效地平衡了用户需求和系统运营之间的矛盾^[24]。科学工作流时间控制所要考虑的另一个问题是工作流时间序列管理问题。工作流时序和时间验证一般需要在工作流执行的一个周期内进行有效的预报,但是现在大多数已有的解决策略都不能很好地处理有限的样品容量和频繁的转折点。为了解决这个问题,Liu Xiao等人提出了一种基于时间序列的预报方法。该方法利用一种周期抽样检验方案建立了有代表性的周期序列,执行时间序列分割,以发现最小模型集,并根据模型匹配结果预测活动周期间隔^[25]。基于这种方法设计的分割算法在预测活动周期等方面显示了很好的效果。

2.7 科学工作流体系结构设计

体系结构设计是指导科学工作流架构的指南,虽然目前已经开发出了多个科学工作流系统,但是对其体系结构的研究相对较少。Lin Cui等人通过对科学工作流系统需求进行详尽分析,设计了科学工作流系统参考模型^[14]。他们认为科学工作流关键的体系结构需求应该包括:支持用户交互和界面定制;支持科研信息的重复利用和演化;能够进行异构服务和软件工具集成;支持异构数据、信息资源的管理;支持高性能计算;能够进行工作流监测和异常处理;支持互操作。从这些需求出发,Lin Cui等人设计的参考模型共分为4层,分别为:操作层,主要管理各种接口;任务管理层,包括任务管理器、来源管理器、数据产品管理等;工作流管理层,主要包括核心工作流引擎、工作流监测器和其他工作流引擎等;揭示层,包括工作流设计、可视化等。

3 结 语

以上的研究介绍反映了当前科学工作流应用过程中亟待解决的关键问题,但是科学工作流在通向成熟的应用、推广过程中需要解决的问题不仅仅是这些。国外学者对科学工作流面临的形势和发展趋势进行了分析研究,归纳这些观点与现状分析,笔者认为,今后科学工作流研究应重点解决以下几个方面的问题:(1)重视科学工作流规范、标准研究;(2)深化科学工作流多系统集成和协作的研究;(3)科学工作流系统应更多解决具体领域的适用性问题;(4)加强使用简单、过程安全的科学工作流解决方案的研究。

相对于国外来说,国内的科学工作流相关应用尚处于初级阶段,其中最主要的原因是当前技术、环境等方面的限制使科学工作流系统的优势没有充分地得以展现。对于图书情报机构尤其是服务于科学研究的图书情报机构来说,紧密跟踪科学工作流的发展状况和发展趋势,建立融入科研过程和科学工作流环境的信息组织、管理和服务能力,也许是数字化环境下图书情报机构的新业务领域。

参考文献：

- [1] Taylor I J, Deelman E, Gannon D B. Workflows for E-Science[M]. New York: Springer, 2007.
- [2] Barga R, Jackson J, Araujo N, et al. The Trident Scientific Workflow Workbench[C]//In Proceedings of the 2008 Fourth IEEE International Conference on EScience. Washington DC: IEEE Computer Society, 2008:317-318.
- [3] Tuot C J, Sintek M, Dengel A R. IVIP: A Scientific Workflow System to Support Experts in Spatial Planning of Crop Production[C]//In Proceedings of the 20th International Conference on Scientific and Statistical Database Management. Berlin: Springer, 2008:586-591.
- [4] Shi Xinghua, Stevens R. SWARM: A Scientific Workflow for Supporting Bayesian Approaches to Improve Metabolic Models[C]//In Proceedings of the 6th International Workshop on Challenges of Large Applications in Distributed Environments. New York: ACM, 2008:25-34.
- [5] Jones M B, Jones C S, Barseghian D, et al. Managing and Analyzing Cabled-Seafloor Sensor Data Using Kepler Scientific Workflows[C]//In Proceedings of the 2008 Ocean Sciences Meeting: From the Watershed to the Global Ocean. Washington DC: American Geophysical Union, 2008.
- [6] Podhorszki N, Ludaescher B, Klasky S A. Workflow Automation for Processing Plasma Fusion Simulation Data[C]//In Proceedings of the 2nd Workshop on Workflows in Support of Large-Scale Science. New York: ACM, 2007:35-44.
- [7] Mandal N, Deelman E, Mehta G, et al. Integrating Existing Scientific Workflow Systems: the Kepler/Pegasus Example[C]//In Proceedings of the 2nd Workshop on Workflows in Support of Large-Scale Science. New York: ACM, 2007:21-28.
- [8] Wei Tan, Foster I, Madduri R. Combining the Power of Taverna and caGrid: Scientific Workflows That Enable Web-Scale Collaboration [J]. IEEE Internet Computing, 2008(6):61-68.
- [9] Gadelha J, Luiz M R, Mattoso M. Kairos: An Architecture for Securing Authorship and Temporal Information of Provenance Data in Grid-Enabled Workflow Management Systems[C]//In Proceedings of the 2008 Fourth IEEE International Conference on EScience. Washington DC: IEEE Computer Society, 2008:597-602.
- [10] Cohen-Boulakia S, Biton O, Cohen S, et al. Addressing the Provenance Challenge Using ZOOM[G]//Concurrency and Computation: Practice and Experience, 2008:497-506.
- [11] Belhajjame K, Wolstencroft K, Corcho O, et al. Metadata Management in the Taverna Workflow System[C]//In Proceedings of the 2008 Eighth IEEE international Symposium on Cluster Computing and the Grid. Washington DC: IEEE Computer Society, 2008:651-656.
- [12] Chebotko A, Fei Xubo, Lu Shiyong, et al. Scientific Workflow Provenance Metadata Management Using an RDBMS-based RDF Store[EB/OL]. [2009-02-26]. <http://www.cs.wayne.edu/~artem/main/research/TR-DB-092007-CFLF.pdf>.
- [13] Wassermann B, Emmerich W, Emmerich W, et al. Sedna: A BPEL-Based Environment for Visual Scientific Workflow Modeling [G]//Workflows for e-Science. London: Springer, 2008:428-442.
- [14] Lin Cui, Lu Shiyong, Lai Zhaoqiang, et al. Service-Oriented Architecture for VIEW: A Visual Scientific Workflow Management System[C]//In Proceedings of the 2008 IEEE International Conference on Services Computing. Washington DC: IEEE Computer Society, 2008:335-342.
- [15] BPEL[EB/OL]. [2009-02-26]. <http://zh.wikipedia.org/wiki/BPEL>.
- [16] Russell N, Hofstede A T, Aalst W V D. NewYAWL: Specifying a Workflow Reference Language Using Coloured Petri Nets[EB/OL]. (2007-10)[2009-02-26]. <http://www.yawl-system.com/documents/newYAWL-cpn.pdf>.
- [17] XPDL[EB/OL]. [2009-02-26]. <http://www.wfmc.org/xpdl.html>.
- [18] Goodman D J. Introduction and Evaluation of Martlet: A Scientific Workflow Language for Abstracted Parallelisation[C]//In Proceedings of the 16th International Conference on World Wide Web. New York: ACM, 2007:983-992.
- [19] Hidders J, Kwasnikowska N, Sroka J, et al. DFL: A Dataflow Language Based on Petri Nets and Nested Relational Calculus[J]. Information Systems, 2008(3):261-284.
- [20] Du Nanshan, Liang Yiwen, Li Qing. Transition System Semantics of DFL as a Scientific Workflow Language[C]//In Proceedings of the 2008 Fourth International Conference on Semantics, Knowledge and Grid. Washington DC: IEEE Computer Society, 2008:241-246.
- [21] Tavares T, Teodoro G, Kurc T, et al. An Efficient and Reliable Scientific Workflow System[C]//In Proceedings of the Seventh IEEE International Symposium on Cluster Computing and the Grid. Washington DC: IEEE Computer Society, 2007:445-452.
- [22] Tolosana-Calasan R, Bañares J A, Álvarez P, et al. Exception Handling Patterns for Hierarchical Scientific Workflows[C]//In Proceedings of the 6th International Workshop on Middleware For Grid Computing. New York: ACM, 2008:1-6.
- [23] Yang Ping, Yang Zijiang, Lu Shiyong. Formal Modeling and Analysis of Scientific Workflows Using Hierarchical State Machines[C]//In Proceedings of the Third IEEE International Conference on e-Science and Grid Computing. Washington DC: IEEE Computer Society, 2007:619-626.
- [24] Liu Xiao, Chen Jinjun, Yang Yun. A Probabilistic Strategy for Setting Temporal Constraints in Scientific Workflows[C]//In Proceedings of the 6th International Conference on Business Process Management. Berlin: Springer-Verlag, 2008:180-195.
- [25] Liu Xiao, Chen Jinjun, Liu Ke, et al. Forecasting Duration Intervals of Scientific Workflow Activities Based on Time-Series Patterns [C]//In Proceedings of the 2008 Fourth IEEE International Conference on EScience. Washington DC: IEEE Computer Society, 2008:23-30.

[作者简介]

苏明明 男，1984年生，中国科学院国家科学图书馆硕士研究生，已发表论文2篇。

宋文女，1961年生，中国科学院国家科学图书馆资源发展部副主任，硕士生导师，已发表论文多篇。

[收稿日期：2009-05-17]