

网格环境下图书馆信息服务探究

□ 黄国彬 / 中国科学院文献情报中心 北京 100080

摘要: 文章分析网格环境下用户信息需求的发展趋势以及这一环境下图书馆信息服务的定位与功能特征, 指出网格环境下的图书馆信息服务将以支持用户创建基于网格技术的PDL为主, 并辅以规则使用的用户指导。在此基础上, 作者结合以网格技术为基础建立起来的PDL原型项目DILIGENT, 剖析了DILIGENT的设计思路、体系结构、主要组件, 最后指出网格环境下图书馆以PDL为支撑的信息服务所面临的挑战。

关键词: 网格, 信息服务, 图书馆

1 引言

随着网格技术的不断发展, 人们已经在建构未来基于网格虚拟环境(e-Science)的科学研究。不过, “e-Science的真面目究竟是什么”却依然是一个令人难以清楚回答的问题。尽管如此, 图书馆的信息服务将继续在网格环境中发挥重要作用, 而且将发生变化却是一个不争的事实。如何保证网格环境下科学家可以足不出户地从图书馆中获得最有价值的各类研究资源已经摆上了图书馆的议事日程。目前, 国内外图书馆界正在不断探索并尝试提供泛在的、个性化的、嵌入式的、动态的、智能化的信息服务, 以使得用户无论是位处于野外台站或者是在虚拟社区, 都能够获得最为贴切和及时的信息服

2 网格环境下用户信息需求的发展趋势

2.1 科研不仅需要文献, 更需要数据

科技文献与科技数据向来是科研的重要支撑资源。一般地, 科学数据可以分为原始的科学研究数据和经过计算加工后的科学数据。前者被人们获取的速度要快于后者以及科技文献被人们获取的速度。在网格环境下, 可供实时共享的资源已经突破了传统的数字化科技文献资源, 更包括各类数据资源、计算资源、仪器资源和设备资源等。科研合作的更加分布化、实时化和大规模化, 不仅要求硬件设备

可以实时高效共享, 更需要各类原始数据的实时共享。但是, 在未经处理的情况下, 原始的传感器数据、实验数据、统计数据并没有太大的实际意义, 可利用性相对较低。因此, 科研人员需要访问必要的计算资源、存储资源和软件环境, 以便对各类复杂的原始数据进行处理。快速提供符合用户需求的衍生数据将在一定程度上成为网格环境下图书馆信息服务的一个重要生长点。

2.2 开放获取与共享

开放获取与共享的需求日益迫切。自存档(Self-archiving)使得科学家个人及其所处的组织(包括虚拟组织)能够发布研究成果并使之供其他科学家共享; 科研组织(包括虚拟组织)可以利用自身创建的规则共享本组织的科研成果(共享并不意味着任何人都可以访问, 而是特定范围内的共享); 科研组织必须拥有必需的资源和技术以便对各类复杂的异构对象进行自存档(存储、检索、访问、掌管、保存); 制定相关机制, 并由科研组织负责注册和发现可获取的自存档资源及其访问规则; 共享规则 and 政策的合理制定及有效实施更是推动开放获取与共享进一步发展的的重要举措。

2.3 跨界知识开发(Cross-domain knowledge Exploitation)

跨界知识开发是网格环境下科研发展的新趋势。一般而言, 科研成果的发布基于特定的环境

(比如,命名、元数据格式、本体、主题词或语言),而跨界知识的全面开发则需要透明访问的支持,而透明访问则需要背景信息(Profile)提供支持;背景信息有利于对被访问的各类知识进行详细解释,它必须是可被共享、维护和保存的知识。在跨界知识开发的过程中,一位科研人员在他人在他人科研成果上进行的各类知识开发,其顺利实现与否取决于这些作为开发源泉的科研成果的质量高低及与之相关的所有服务是否得到全面实现。此外,跨界知识开发更需要有稳定可靠的共享服务,即共享必须具有可获取性、广泛性和高效性。

3 网格环境下图书馆信息服务的定位与功能特征

3.1 网格环境下图书馆信息服务的定位

针对图书馆在网格技术环境下的信息服务定位,国外不少学者对此展开了研究。其中尤以学者 Paul Miller 的观点最具有代表性^[1]。他指出,在图书馆界,我们已经意识到更广阔的信息空间中所出现的各种发展趋势。同时,我们也意识到,人们对于获取信息和服务的需求在持续上升。人们希望改变传统面对面的服务提供方式,更希望通过网页以及各种可以实现电子化的通讯手段来获取服务。越来越多的人很愿意共享数据、深度挖掘数据背后的信息以及通过集成数据源获取到更多更有价值的信息。此外,我们也已经感受到,图书馆的用户、馆员、图书馆的资助者、数字图书馆的研究者以及各方人士将希望能够利用现在或者是以后可以将他们紧紧结合到一起的系统,使用更为广泛的方式与手段,突破传统围墙的阻隔和现有部分系统不兼容的限制,更好地实现网格环境下图书馆自身的价值。

2006年,Paul Miller又在其另一篇著作中阐述了他关于library 2.0环境下图书馆信息服务的观点^[2]。他指出,显而易见,为了实现library 2.0的各种目标,使得图书馆成为一个更加可以被感受和被利用的、能够提供价值含量高的信息资源,图书馆在各个方面都必须进行改革。作为图书馆事业的从业人士,我们必须打破信息单元(Silos)之间的壁垒、敲开系统之间的隔阂、跨越团队之间的封闭。我们需要在图书馆与用户之间构建、培育并引领起一个真正

的具有可持续发展的社区。而这一切均根源于,我们今后需要共享比今天更多更为复杂的内容。我们需要向开源理念的践行者吸取经验,我们也要向推动网络不断创新的社区发展规范(Community-nurtured specifications)这一理念取经,突破传统管理方式中存在的思维定势。

事实上,用户环境的进一步网络化、虚拟化,要求图书馆的信息服务无论在内容、方式、手段等方面都应该网络化、虚拟化。因此,具有高度虚拟化信息服务特征的PDL(Personal Digital Library)将成为图书馆在网格环境下的主导信息服务媒介。当然,既然复合图书馆将存在一定的时间,作为复合图书馆信息服务存在形态的具体表现形式,虚拟服务与面对面的阵地服务也依然将并存一定的时间。尽管目前图书馆仍然会为用户提供到馆服务(in person),甚至“上门服务”,如中国科学院国家科学图书馆更是提供以学科馆员为保障的“到所服务”。实际上,随着用户素养的提高以及网格共享技术的深入发展,这些服务都将逐渐成为图书馆辅助性的信息服务。资源并不是图书馆核心竞争力的体现,图书馆的核心竞争力应体现在服务上^[3]。学科化服务是不是图书馆核心能力的体现?是否是未来图书馆信息服务的主流?这些都是有待实践进一步验证的问题。

尽管在我国目前仍未大范围地将网格技术应用于数字图书馆建设,故网格为数字图书馆带来的实质性收益仍未为人们所感受到。事实上,包括美国、欧盟在内的多数发达国家和地区已经在2004年就投入巨额资金用于研发基于网格技术的数字图书馆,相关的项目包括OpenDLib、COLLATE、BRICKS、DILIGENT、SRB、CheshireIII、ARCO、GRACE、DELOS、Digital Library Grid、FreeLib、GridIR等。纵观这些项目的立项动机,无疑是要在网格环境下进一步提高图书馆以数字图书馆为服务途径与渠道的信息服务能力。有鉴于此,在复合图书馆依然存在一定时间这一前提下,图书馆一方面需要加强传统的面对面服务,更应该着眼于未来,敢于引进、吸收先进的网格技术,强化自身的虚拟服务能力。

通过分析国外基于网格技术建立起来的各个数字图书馆原型项目的发展目标,笔者认为,网

格环境下,图书馆信息服务更应该定位于:支持用户创建属于自己的真正PDL,而不局限于现在如Mylibrary这种功能相对单一的、服务相对静止的数字图书馆服务,要使得数字图书馆成为图书馆在网格环境下的一种泛在服务(Ubiquitous Service),同时为用户提供PDL标准与规则的制定与使用方面的指导,保证用户可以高效率地应用各类PDL服务。

3.2 网格环境下图书馆PDL信息服务的功能特征

网格环境下,PDL将包含当前个性化数字图书馆的全部功能。同时,也具有自身与网格技术相结合后的各类新功能。这些新功能体现为:高效管理各类与虚拟组织开放研究密切相关的信息资源;有效支持信息资源生产的整个生命周期,允许以合作的方式管理信息资源的创建、浏览和相关操作;能够按照信息资源提供者制定的规则,为虚拟组织提供一种可控制的、个性化的信息资源共享机制;根据虚拟组织内部的需求与面临的外部环境的变化,及时调整适应,并提升自身的个性化服务能力。

4 面向网格环境的PDL原型项目介绍

目前,以网格技术为底层技术支撑,期望提供泛在的、个性化的、嵌入式的、动态的、智能化的信息服务的数字图书馆项目发展迅速。较有代表性的项目是DILIGENT项目,该项目由欧盟资助,并由意大利国家委员会所属的信息科技研究所承办。下文将以DILIGENT为例,分析基于网格的PDL的设计思路、体系结构及主要组件的功能。

4.1 DILIGENT原型系统的设计思路

DILIGENT项目的建设者认为,以网格技术为支撑的PDL,必须在代理(Surrogate)、虚拟组织(Virtual Organization)、可定制性(Customized)以及动态调度(Dynamic Dispatch)上给予充分考虑^[4]。

4.1.1 构筑数字图书馆的代理

从抽象的理论层面来分析,DILIGENT开发的基础体系架构实质是一种数字图书馆的代理,而这一代理沟通的客户双方分别是数字图书馆信息资源的提供者和信息资源的消费者。信息资源提供者一

般是个人或者某个组织,他们在遵循DILIGENT基础体系架构的条件下,设计出特定的资源访问入口和访问政策,将自身拥有的资源予以发布,提供其他个人或虚拟研究组织共享。资源建设者则是希望建立PDL的用户群。一般情况下,由DILIGENT基础体系结构作为代理进行管理的资源包括:各类资源集,这类资源主要是通过单个访问点可以查找和访问到的信息对象集;服务,主要是具有某种特殊功能的软件工具,而且其描述、界面和使用限制已经得到明晰界定并且是可供公众获取;服务节点,一般是提供计算和存储能力的网络实体,它的功能集中在于为资源收藏和服务提供一个活动环境(an environment for hosting collections and services)。

4.1.2 以虚拟研究组织的组成机制为基础

为了支持资源提供者和资源消费者之间可控制的资源共享,DILIGENT基础体系架构的设计完全依赖于已经在网格领域被广泛引用的虚拟研究组织机制。这种机制主要是以各种高级的共享控制规则将用户和资源集汇总在一起并建立模型。通常情况下,它是建立在一种证明基础框架之上的。虚拟研究组织具有受限制的生命周期,而且可以被动态创建,以满足各种特定的需求,保证资源的分配和提供完全可因需求的变动而变化。通于开发由DILIGENT基础体系结构提供的合适机制,信息资源提供者可以对其提供的各类资源进行描述以便实现资源注册。根据其所提供的不同资源类型,DILIGENT的基础体系结构能够自动地从资源中抽取资源提供者未使用的其他属性及属性值来描述资源,以便进一步丰富该信息资源的显性描述信息。随后,DILIGENT基础体系结构将为各类已经注册过的资源进行管理,包括各种发现、监控和保存操作,并依据这些资源执行某些相关功能,从而为各种服务质量和可控制的资源共享提供某种程度上的支持。

4.1.3 提供可定制的数字图书馆服务

虚拟研究组织的用户群可定义一组具体的条件集,并对各种可获取的资源进行适当的组配与汇总,构建出一个或者是多个适合自身需求的数字图书馆。具体条件集的确定可根据各个信息空间(Information Space,主要包括信息资源收藏集、信息资源内容的主题以及文档类型)、服务质量

(一般是指可获取性、性能、安全等)以及其他各种因素(比如最大的成本、最长的生命周期)的不同而有所区别。数字图书馆代理将从虚拟研究组织的用户群所提出的数字图书馆构建条件出发,在可获取的范围内将符合需求的各类资源加以选择(多数情况下,DILIGENT基础体系结构也会对各类资源进行部署),并采取合理的策略将这些资源进行汇总。最后,以门户的形式将新创建的PDL提供给虚拟研究组织的用户,使其可以访问根据其提出的收集策略与条件而汇总、集成在一起各类资源。

4.1.4 动态调度各类可支配的信息资源

DILIGENT基础体系结构将持续对其可以管辖的各类数字图书馆所拥有的各类资源进行持续监控,如果有需要,该基础体系结构甚至会对这些资源进行调整与变动,使相关数字图书馆汇集的资源可以提供更高质量的服务。正因为这样,数字图书馆的组合与联盟往往是一个动态的过程。这在一定程度上也使得PDL的创建与修改可以在极其短暂的时间内完成,并且可以由创建它的虚拟研究组织来执行相应操作。与以往相比,用户在数字图书馆的创建与修改,再也不需要考虑大量的资金投入,更不需要投入大规模的建设力量。

4.2 DILIGENT的体系结构

下文将从信息空间管理区、访问区、表示区、用户与资源空间管理区、功能实现支持区等主要组件剖析DILIGENT的体系结构^[5]。

4.2.1 信息空间管理区

信息空间管理区(Information Space Management)包含有内容管理服务。内容管理服务主要是提供复制管理和存储管理,并对DILIGENT信息对象存储库进行揭示。其次,元数据管理(Metadata Management)服务借助内容管理功能模块提供的服务功能,对元数据的存储与管理加以控制;第三,提供注解管理功能的注解服务模块将把自身附载到信息对象上;最后,内容安全管理服务模块将通过采用各种数字水印技术和加密技术,保证各个虚拟研究组织自行创建PDL所存储与管理的数字对象不被未经授权的其他虚拟研究组织、个人或者用户群使用。在DILIGENT项目开发的基础体

系结构中,存储管理完全依赖于网格存储设备,其存储过程就是将物理文档存放到不受DILIGENT基础体系结构直接控制的第三方设备上,正因为这样,内容安全管理服务模块的功能能否正常发挥显得至关重要。

4.2.2 访问区

访问区包括有查找服务,其功能的正常发挥依赖于DILIGENT基础体系结构中其他多个组件的支撑,这些组件包括:索引管理(Index Management),它包含有多个组件,而且这些组件能够对信息领域中各种类型的索引进行创建和维护,比如,它可以对全文、索引等索引进行创建与维护;特征抽取(Feature Extraction),它是收集各种类型的特征抽取组件,而这些组件往往因媒体类型的不同而有所不同;内容来源描述与选择(Content Source Description and Selection);数据融合(Data Fusion),它是一种能够对来自各个不同收藏集的当前结果集进行管理和控制的组件;个性化,即根据用户的使用偏好与检索兴趣点,提供用户将检索结果进行个性化定制的组件。所有的这些服务组件的联合发挥作用,必须依靠过程管理组件(Process Management)提供支撑。总而言之,若要对各种请求作出快速响应,则需要查找服务模块制定出一个可执行计划,而且这种计划必须以用户请求为基础,并将这些请求以工作流的方式建模,建模完成后,请求模型将由执行服务(Execution Service)以P2P的方式进行处理。

4.2.3 用户和资源空间管理区

用户和资源域管理区提供动态虚拟组织支撑服务。事实上,这种支撑服务是一种“元服务”(Meta-service),它负责为虚拟组织在资源处理方面的活动开展提供支持,而虚拟组织则通过创建一个可信任的环境,确保DILIGENT各项资源共享得到有效控制。为了实现这一任务目标,“元服务”需要提供一系列的服务,也需要其他一系列的服务提供支持:通知服务(Notification Service)将负责把特定主题的信息通知给用户,比如,可获取的各类新资源的推介;用户和群体管理服务(User and Group Management)将为用户注册和管理任务提供支持,并且可以为群体内部的用户组织提供相

关机制；资源注册支持服务（Resource Registration Support Services）则允许用户将新出现的资源添加到DILIGENT基础体系架构中。

4.2.4 表示区

表示区（Presentation area）完全面向用户，并且从用户的需求出发。它能够自动生成面向特别用户群的访问门户（Portals），从而可以为用户提供数字图书馆个性化的访问服务。特别需要指出的是，这种访问门户能够支持各类插件，也允许用户群在访问门户时运行各种可视化工具。从技术的角度分析，DILIGENT使用一种开源的网站—主机（Portlet-hosting）引擎作为搜索支持，该搜索引擎遵循JSR1683、JavaServer Faces 4和WSRP5等标准规范，对于作为DILIGENT服务一个完整组成部分的用户界面，该搜索引擎也可以进行集成。

4.2.5 功能实现支持区

功能实现支持区将主要提供以下功能：由信息服务功能模块（Information Service）监控和发现DILIGENT系统内部所有可获取的资源；采用全局化战略，保证用户可以最优化地利用DILIGENT基础体系结构提供的各种主机节点资源；协调各类用于构建虚拟数字图书馆的资源池（Resource Pool），并由守护者服务（Keeper Service）来保证系统具有一定程度的容错能力；支持工作流的设计与认证，通过作业管理服务确定工作流设计与认证的可执行性和最优化。

5 网络环境下图书馆以PDL为支撑的信息服务所面临的挑战剖析

尽管网络环境能够为图书馆创建真正的PDL带来技术上的便利。但是，网络自身的优点，如分布性，却给数字图书馆资源的高效管理带来一定的挑战。此外，PDL创建的自发性、数据管理的自发性以及海量数据之间存在的异构性，也在很大程度上给PDL带来发展风险^[6]。

5.1 分布性方面面临的挑战

在网络环境下，分布性体现在数据分布和过程分布两方面。由于数字图书馆管理的写入数据、内容数据以及用来标引和分类数字数据的原数据全部

分布在网络中不同的节点上，这种分布式的结构既是由于版权问题所必须采取的方式，也是为了解决某一单个节点上无法对如此大量数据进行管理的问题。在这种情况下，数字图书馆中的一些重要操作（例如更新数据/元数据等需保持数据连续性的操作）就变得更加复杂了。虽然分布式使更多数据的处理成为可能，但是如何在这样一个分布式数据集合中进行联邦检索就成为一个不得不面对的问题。伴随着数据分布，检索和其他数据处理过程也必须实现分布式处理。虽然分布式检索和结果融合非常容易被理解，但是如何安排每一过程执行的进程，以减少通信开销仍是目前无法圆满解决的问题。

5.2 自发性方面面临的挑战

网络不同于簇技术（Cluster），参与的节点不是单一的并且可能不能以一种可靠的方式获取。但是，对于一个数字图书馆，数据需要永久的可获取，因为数字图书馆的功能之一就是长期的数据存储。在一个可能不可靠的系统之上构建一个要求数据可靠的系统，这就要求大量的数据复制。来自许多节点的数据被复制以确保即使某些节点的数据不可获取的时候，至少有一份数据是可靠的。但是，可读写系统的复制在语义更新方面有其自身的复杂性。如何在有些复制的数据缺失的情况下更新数据，或者是，在不同的复制数据被不同的用户进行更新的情况下，如何确保被复制数据的最新一致性也是一项严峻的挑战。

5.3 异构性方面面临的挑战

较之集中式的环境，分布式环境下的元数据和数据异构的可能性更大。在数字图书馆环境下，异构的各种资源集合可能包含不同类型的数据（例如，天气报告、考古发现视频影像）。再者，每一种集合本身也包含各类异构数据，由于数字图书馆数据容量的限制，可能需要将数据分布为许多交织的结点进行存储。目前，数据的异构性主要包括：句法异构和语义异构。句法异构包括数据格式、数据模型；语义异构包括naming和半自动拓扑。如何将异构的资源格式化为统一的XML进行处理也是网络环境下PDL建设必须面对的难题。

（下转77页）

- [7] SHNEIDERMAN B. 用户界面设计—有效的人机交互策略[M]. 张国印, 等译. 北京: 电子工业出版社, 2004: 3.
- [8] 肖婷. 美国少儿数字图书馆初探[J]. 江西图书馆学刊, 2006(2): 102-103.
- [9] 中国少儿数字图书馆. 中国少儿数字图书馆简介[EB/OL]. [2007-06-16]. <http://www.cclib.com.cn/about/>.
- [10] ICDL. International Children's Digital Library[EB/OL]. [2007-06-16]. <http://www.childrenslibrary.org/>.
- [11] ICDL. International Children's Digital Library[EB/OL]. [2007-06-16]. <http://www.childrenslibrary.org/about/fastfacts.shtml>.
- [12] 曹玲. 国际儿童数字图书馆 (ICDL) 建设特点及其启示[J]. 图书馆杂志, 2006(3): 65-67.

作者简介

刘敏 (1982-), 研究方向为信息资源管理. 通讯地址: 中山大学新港西路135号园东区105#711 510275

A Research about the User Interface of the Children's Digital Library

Liu Min / Department of Information Management, Sun Yat-Sen University, Guangzhou, 510275

Abstract: This paper first briefly introduces the concepts, evaluation principles and indexes of the Children's Digital Library, then takes the International Children's Digital Library-ICDL and the Chinese Children's Digital library as examples to analyze the advantages and disadvantages of their user interface.

Keywords: Children's digital library, User interface, Availability

(收稿日期: 2007-11-20; 责任编辑: 虞敏)

(上接63页)

6 结语

科学研究日益成为一项协作性和跨领域特别强的工作。随着网格等信息技术的发展, 图书馆信息服务在网格环境下发生变化已是一种必然。为了更好地支持虚拟组织的研究需求, 图书馆在网格环境

下提供的信息服务, 将不再基于一个静态的预先定义的信息资源获取系统, 而应能够根据虚拟组织需求的动态变化, 为虚拟组织成员提供灵活可定制的各类资源与服务, 并支持生产、开发、传播和共享各类复杂的资源对象。

参考文献

- [1] MILLER P. Coming Together around Library 2.0: A Focus for Discussion and a Call to Arms[J/OL]. 2006-04[2007-09-19]. <http://www.dlib.org/dlib/april06/miller/04miller.html>.
- [2] Miller, P. Library 2.0: The challenge of disruptive innovation[EB/OL]. 2006[2007-07-19]. http://www.talis.com/resources/documents/447_library_2_prf1.pdf. 2007-07-19.
- [3] 初景利. 博士生课程“用户研究与信息服务”授课讲义. 2007-07-18.
- [4] DILIGENT Project Web Site, Digital Library Infrastructure on Grid Enabled, Technology[EB/OL]. [2007-09-01]. http://dlib.isti.cnr.it/pub/bscw.cgi/d17909/D4.2.1_Website.
- [5] CANDELA, CASTELLI D, LANGGUTH C, et al. On-Demand Service Deployment and Process Support in e-Science DLs: the DILIGENT Experience[C/OL]. Proceedings of the International Workshop on Digital Library Goes e-Science. 2006-09.
- [6] MEHTA B and FANKHAUSER P. To Grid or not to Grid: Digital libraries based on Grid infrastructure[EB/OL]. [2007-09-19]. <http://www.l3s.de/~mehta/ecdl.pdf>.

作者简介

黄国彬, 中国科学院文献情报中心博士生. 通讯地址: 北京中关村北四环西路北一街33号 中国科学院文献情报中心 100080

Exploration on Library Information Service Under Grid Environment

Huang Guobin / Documentation and Information Center of CAS, Beijing, 100080

Abstract: The paper analyzes the trend of information need, the position and function features of library information service under grid environment, pointing out that library information service will focus on supporting users to create PDL based on grid technology and give users rules instruction. On that basis, it takes PDL prototype DILIGENT for example and discusses the design ideas, system structure and main components of DILIGENT. Finally, it explores the challenges of library information service based on PDL under grid environment.

Keywords: Grid, Information service, Library

(收稿日期: 2007-11-05; 责任编辑: 虞敏)