

·信息工作·

# 新技术思想与数字图书馆发展研究

奉国和 (华南师范大学经济管理学院 广东广州 510006)

**摘要** 文章通过探讨网格、对等网、面向服务、云计算、数据挖掘、软件即服务、开源软件、Library2.0等新技术思想原理及在数字图书馆中的应用,分析各种技术思想之间的关联关系,指出基于新技术思想的数字图书馆建设应注意问题。

**关键词** 数字图书馆 云计算 SaaS SOA 开源软件

中图分类号:G250.76

文献标识码:A

文章编号:1003-6938(2010)02-069-05

## Research on Digital Library Based on New Information Technology

Feng Guohe (School of Economy & Management, South China Normal University, Guangzhou, Guangdong 510006)

**Abstract** : This paper discusses the new technology theory and application in digital library such as Grid, P2P, SOA, cloud computing, data mining, SaaS, OSS and Library2.0 and analyzes the associate relationship between these new technologies. Then, points out some issues which are needed paying attention to digital library construction based on these new technologies.

**Key words** : digital library ; cloud computing ; SaaS ; SOA ; open source software

CLC number: G250.76

Document code: A

Article ID: 1003-6938(2010)02-069-05

经济、社会环境等因素是数字图书馆发展的重要基础,而技术因素则是数字图书馆发展的直接推力。IAN ROWLANDS等给出的数字图书馆建设的生命周期模型指出图书馆引进、探索、整合、传播、使用等业务工作都是在经济、社会、信息等因素影响下进行的。<sup>[1]</sup>近年来,新技术思想的出现对数字图书馆建设产生了深远的影响,改变着数字图书馆发展的方向,特别是2004年6月在意大利召开的“数字图书馆的体系结构:对等网、网格和面向服务”会议、Library2.0以及Michael Stephens预测2009年对图书馆有影响的十大技术趋势之“无所不在的云计算”、“信息技术的角色变迁”等让很多图书馆建设者重新思考图书馆的架构、管理与服务模式。<sup>[2][3]</sup>本文对当前出现的新技术思想进行了研究,并就其如何在数字图书馆建设中应用作了分析,可以预见新技术对数字图书馆的影响是深远的。图书馆建设者在实际工作中多了解新技术新思想有利于工作的开展,推动数字图书馆为人们、为社会提供优质服

### 1 新技术思想原理与应用

#### 1.1 新技术分析

##### (1) 网格、对等网、SOA技术

2004年6月在意大利召开的“数字图书馆的体系结构:对等网、网格和面向服务”会议中,<sup>[5]</sup>学者围绕网格(grid)、对等网(peer to peer network, P2P)与面向服务(service-oriented architecture, SOA)等技术讨论构建数字图书馆新体系结构问题。Ingo Frommholz指出网格、对等网与面向服务等技术构建下一代数字图书馆体系结构成为可能。(1)设计高效低成本的下一代数字图书馆体系结构,使数字图书馆技术向更广泛的用户群体开放。(2)快速适应DL服务及相关IT领域的发展。(3)促使更广泛的内容与服务提供者的动态联邦模式。<sup>[6]</sup>Walkerline J等提出了一种基于P2P的数字图书馆系统结构原型(见图1)。<sup>[7]</sup>吴颖红指出图书馆利用P2P技术可以进行文件交换、对等计算、协同工作、即时通讯、信息搜集等工作。<sup>[8]</sup>这些工作也是图书馆日常进行的基础工作,利用P2P技术可以给图书馆工作带来便利。网格在我国的典型应用为中国国家网格(China National Grid, CNGrid),CNGrid是国家863计划“高性能计算机及其核心软件”重大专项的重要组成部分,通过资源共享、协同工作和服务机制,有效支持科学研究、资源环境、先进制造和信息服务等应用。<sup>[9]</sup>

##### (2) 云计算

云计算(Cloud Computing)是分布式处理、并行处理

基金项目:本文系国家社科基金项目(编号:08CTQ003)与广东省哲学社会科学规划项目(编号:06M03)研究成果之一。

收稿日期:2009-02-10 责任编辑:刘全根

和网格计算的发展,基本原理为通过使计算分布在大量的分布式计算机上,而非本地计算机或远程服务器中,企业数据中心的运行将更与互联网相似。这使得企业能够将资源切换到需要的应用上,根据需求访问计算机和存储系统。充分体现了以用户为中心的理念,并具有超强的计算能力。李永先指出云计算将给图书馆工作带来如下深刻变化:①确保服务器的可靠运行,最大限度地降低服务器出错概率;②以极低的成本投入获得极高的运算能力,克服服务器访问限制的瓶颈;③在更大程度上进行信息资源共享;④降低有关的维护费用;⑤最大限度地发挥图书馆的作用。<sup>[10]</sup>Michael Stephens预测2009年对图书馆有影响的十大技术趋势时将“无所不在的云计算”列为首位,他认为云计算将会对图书馆产生重大影响。<sup>[11]</sup>云计算所体现的本质特征,也是图书馆需要的,将图书馆架构在云上具有无比的优越性,极大提升图书馆的服务能力。

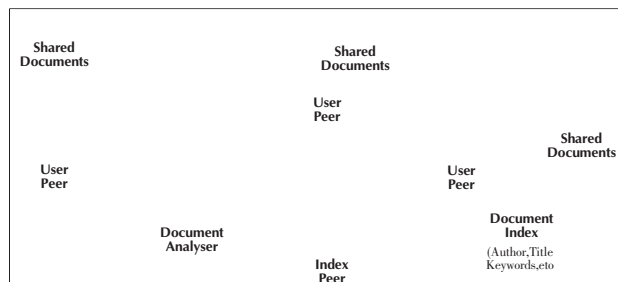


图1 基于P2P的数字图书馆系统结构原型

### (3) 开源软件

开源软件(Open Source Software, OSS)直接字面意思是公开源代码的软件,其经过十几年的发展,已经成功应用到很多行业,为各组织的发展起到相当大的作用。美国的开放源码促进会(Open Source Initiative, OSI)对于开源软件描述非常详细,大致可以分成四个层面,首先是软件具有无条件的发布自由;其次是对于源代码开放的具体说明;第三关于演绎作品,即开放源代码软件许可证必须允许源代码可以被修改和允许根据源代码产生有版权法意义上的演绎作品,而这一点为日后开源软件的商业用途提供了契机;最后是与软件使用、开发是否受到歧视相关的要求,例如不得对许可对象有歧视、不得对软件被使用的领域有歧视。<sup>[12]</sup>开源软件深得用户喜爱,一方面是由开源软件自身的特点决定的;另一方面则是社会与技术的发展决定的,开源软件对于客户而言意味着多样性和替代性,更强的客户化能力以及防止供应商锁定,提高透明性和安全性。而对研究或开发者而言则意味着更快的开发速度,更低的开发成本,更少的重复劳动,快速的错误修正,加速了知识共享与知识创新,这些都是与社会需求相适应的。由于开源软件的这些价值,开源软件正日

益表现出旺盛的生命力,目前有很多的开源软件运用于数字图书馆建设中,也取得了不错效果,但从更广泛的范围来看,它还有很大的发展空间。奉国和指出图书馆在数字化建设中与开源软件结合是一种趋势,利用开源软件不仅可以构建图书馆知识管理系统还可以提升图书馆的创新能力。<sup>[13][14]</sup>钱国富认为,基于开源软件不仅有助于图书馆在有限的技术力量和资金投资下进行图书馆2.0建设,而且还能够拓展图书馆视野,打破商业软件对图书馆技术发展的束缚。<sup>[15]</sup>目前很多开源软件已经应用到数字图书馆建设中,如操作系统Linux等,数据库MySQL、PostgreSQL等,内容管理Dspace、Fedora等,检索引擎Lucene、Nutch等,这些开源软件为加快图书馆建设、节省图书馆经费带来了极大便利。

### (4) 数据挖掘与知识发现

数据挖掘是指从大量的数据中提取出可信、新颖、有效并能被人们所理解的、潜在的模式、规律或趋势的高级处理过程。它融合了数据库、人工智能、机器学习、统计学等多个领域的理论和技术,在理论研究领域也被称为数据库中的知识发现。<sup>[16]</sup>数据挖掘最主要和最重要的分析方法有四种:①关联分析,即利用关联规则进行数据挖掘。关联分析的目的在于分析决定哪些事情将一起发生。如:它能发现数据库中形如“90%的读者在一次借阅活动中借阅书籍A的同时借阅书籍B”之类的知识;②序列模式分析,与关联分析相似,其目的也是为了挖掘数据之间的联系,但序列模式分析的侧重点在于分析数据间的前后序列关系。它能发现数据库中形如“在某一段时间内,读者借阅书籍A,接着借阅书籍B,而后借阅书籍C,即序列A→B→C出现的频度较高”之类的知识;③分类分析,设有一个数据库和一组具有不同特征类别(标记),该数据库中的每一个记录都赋予一个类别的标记,这样的数据库称为示例数据库或训练集。分类分析就是通过分析示例数据库中的数据,为每个类别做出准确的描述或建立分析模型或挖掘出分类规则,然后用这个分类规则对其它数据库中的记录进行分类;④聚类分析,与分类分析不同,聚类分析输入的是一组未分类记录,并且这些记录应分成几类事先也不知道。聚类分析就是通过分析数据库中的记录数据,根据一定的分类规则,合理地划分记录集合,确定每个记录所在类别。<sup>[17]</sup>

图书馆馆藏数量越来越大,而读者感兴趣的只是其中很少一部分,如何在浩瀚的书籍信息中找到自己感兴趣的信息则是一件很头疼的事情,如何帮助读者高效找到所需信息资源是图书馆要研究的问题。图书推荐服务是采取主动的方式,向读者提供他可能感兴趣的信息,降

低读者对专业知识的要求,节约了读者寻找资料的时间,同时提供了更丰富的信息。图书馆通过分析读者借阅行为来分析读者的借阅特征,发现读者的兴趣爱好,从而采取有个性化的服务;通过分析点击流来分析用户的一些行为特征,发现背后隐藏的信息,通过分析图书的借阅情况,发现哪些书是畅销书,哪些书是流通率小的书籍,为图书采购部提供指向。为此需要从读者的借阅行为中发现读者的借阅特征或者习惯,建立相应模型提供推荐服务。数据挖掘可帮助图书馆实现个性化推荐、图书采购指导、读者群分等服务工作,提升图书馆服务能力。

## 1.2 新思想分析

### (1) 软件即服务

软件即服务(Software-as-a-Service, SaaS)是随着互联网技术的发展和应用程序的成熟,在21世纪开始兴起的一种完全创新的软件应用模式。它是一种通过Internet提供软件的模式,厂商将应用程序统一部署在自己的服务器上,客户可以根据自己实际需求,通过互联网向厂商定购所需的应用程序服务,按定购的服务多少和时间长短向厂商支付费用,并通过互联网获得厂商提供的服务。与传统模式比较SaaS具有使用费用低廉、按需求定制、发展潜力空间大等优势。<sup>[18]</sup>SaaS模式对图书馆建设具有极大的启发,基于SaaS模式图书馆可以将一整套的数据库、软硬件系统交付专业公司建设管理与维护,而图书馆只需专注最核心的服务工作,该模式或许是图书馆以后发展的新方向。

### (2) Web2.0、Library2.0理念

很多学者对Web2.0进行过阐述,笔者认为Web2.0是技术跟理念的融合,Web2.0使得网络向社会化和个性化发展,Web2.0的核心理念强调“以用户为中心”,鼓励用户积极参与信息资源的建设。基于Web2.0图书馆界适时提出Library2.0概念,最早出现Library2.0概念在“Campus Library 2.0”<sup>[19]</sup>文中,之后相继有不同的研究者在不同的场合提出Library2.0的概念,对Library2.0作了不同的阐释。Michael Casey在“Library Crunch”中对Lib2.0的解释是“Library2.0是一个能够让图书馆快速地反应市场需求的运作模式,这并不表示要放弃既有的读者或使命。它是一种快速改变、弹性的组织架构、Web2.0工具、及使用者参与的哲学,能够使图书馆处于一个更加巩固的位置,并有效地、有效率地满足更多使用者的需求。”<sup>[20]</sup>John Blyberg指出“Library2.0在馆员之间或者是馆员与读者之间深受科技导向、双向、及社交互动(social interactions)的影响。Library2.0提供了一个框架,在该框架里以usability及findability的目标来重新评估传统图书馆经营的每一个面

向”。<sup>[21]</sup>国内《图书馆杂志》2005年12期有Web2.0对图书馆创新服务的启示论文;<sup>[22]</sup>《大学图书馆学报》2006年1期有关于图书馆2.0的论文;<sup>[23]</sup>而后相继与一些Web2.0/Library2.0的研讨会召开。刘炜将Library2.0定义为:Web2.0的理念和技术在图书馆行业中的应用,它不仅是一些技术的组合,而且是新的“思考方式”或“运营方式”,Lib2.0利用Web2.0技术和理念为图书馆用户提供按需定制的信息服务,如互联网作为平台、利用集体智慧、用户添加数据、支持社会性网络、丰富的用户体验等等,它以用户需求或服务为导向,而不是以技术为导向。<sup>[24]</sup>

从Library2.0的解释来看,Library2.0可以看作Web2.0的延伸,是技术、服务、创新等的结合,Library2.0的提出将会对数字图书馆的发展起着重要作用。基于Web2.0图书馆可以做到:图书馆利用RSS实现信息推送和共享,利用Blog、Wiki建立交流平台,专题信息库等提高图书馆的信息服务水平,利用IM软件促进图书馆的虚拟参考咨询服务建设,利用Tag实现图书馆信息的分类和组织,利用社会网络促进用户与图书馆的合作与共享等。Library2.0的出现极大丰富了图书馆工作内容与工作方式。

### (3) 其他

SOA技术,其实也是一种思想的重要体现,面向服务结构强调以服务为中心架构图书馆体系,服务思想应该在该结构中得到重点体现。另外开源软件也体现出一种思想,该思想强调图书馆建设开源模式的推行,该模式与商业模式不同,颠覆了图书馆传统的建设模式,影响图书馆未来的发展。

## 2 新技术思想关联关系及启示

新技术为图书馆建设带来新思路,数字图书馆建设充分利用新技术可以节约成本,加速建设进程。图书馆根据各种技术思想特点,有针对性的选择适合自己的建设模式。为此,理清各种技术思想的关联关系,找到一种最佳组合模式是运用新技术的前提。

### 2.1 开源软件与 Web2.0/Library2.0

开源软件已经被各行业广为认识和接受。同样在图书馆的建设中,开源软件也受到了图书馆人的认同和欢迎。范并思则认为开源软件是Web2.0的重要理念,应用开源软件也成为Library2.0的重要应用。<sup>[25]</sup>钱国富认为基于开源软件进行Library2.0建设,能够拓展图书馆视野,打破商业软件对图书馆技术发展的束缚。<sup>[26]</sup>目前在全球数字图书馆的研究和实践过程中,出现了很多能够支持图书馆应用的开放源码软件,例如支持存储检索、收割聚合、开放链接、集成门户、知识技术、可视化、单点登录、内容



管理、机构仓储、虚拟学习、学术交流等多个方面的开放源码软件。开放源码思想的兴起和开放源码软件的蓬勃发展,给数字图书馆的发展提供了一个“低代价可持续发展”的机会、一个以高性价比产品获得竞争力的机会。针对数字图书馆的发展情况,如果有效地选择、利用、本地化和再开发目前比较成熟的开放源码软件,对于提高自身数字图书馆的层次和起点,加快自身数字图书馆的发展有着重要的意义。开源很多思想与Web2.0/Library2.0的理念是相同的,它们都主张开放、共享,强调人人参与。

## 2.2 网络与云计算

Grid是构筑在互联网上的一组新兴技术,它将高速互联网、高性能计算机、大型数据库、传感器、远程设备融为一体,为科技人员和普通百姓提供更多的资源、功能和交互性。网络技术利用高速因特网把分布于不同地理位置的计算机、数据库、存储器和软件等资源连成整体,就像一台超级计算机一样为用户提供一体化信息服务,其核心思想是“整个因特网就是一台计算机”。网络目标本身不在于规模的大小,而在于可共享资源的种类,共享使用资源的形式,对用户共享资源的要求,共享的透明程度,接口的简单程度等。网络与云计算有很多相似之处,但是云所包含的概念更广泛。高岚岚对网络与云计算从作业调度、应用影响、服务理念等方面作了深入比较。<sup>[27][28]</sup>可以预见,云计算无比巨大的优越性,必将对未来产生深远的影响。

## 2.3 云计算与 SaaS

云计算的内涵即体现了SaaS思想,它通过浏览器把程序传给成千上万的用户。在用户眼中看来,这样会省去在服务器和软件授权上的开支;从供应商角度来看,这样只需要维持一个程序就够了,这样能够减少成本。云计算核心思想为无所不在的计算、服务,用户只要接上云就可以开展自己工作,而为此提供的服务工作则全部交由云提供商去完成,所以未来将SaaS思想与云计算相结合是一种趋势。

## 2.4 SaaS 与 SOA

SOA是一种框架的方法,它是开发松耦合软件的框架,是描述业务组件如何构架的方法;而SaaS是一种传递模型,是描述软件如何交付和传递的方法,基于SaaS模式的应用程序是指单实例多个承租(Multi-Tenant)的应用软件。SaaS和SOA关注问题的角度不同,都是时下流行且备受推崇的技术,两种技术是共生的,而且,两者可以通过两种不同的方法一起工作。一方面,SaaS应用软件程序需要通过SOA的标准和SOA的方法建立起来并作为标准服务被传递;另一方面,SOA基础设施可作为SaaS应用软

件程序的一个集成点被使用,通过在SaaS应用软件程序和内部业务线应用软件程序之间,应用基于SOA架构的组织服务总线,可以实现应用集成。<sup>[29]</sup>

## 3 应用新技术思想建设数字图书馆应注意的问题

### (1)坚持以服务为宗旨

服务是图书馆的核心工作,任何先进的技术思想都应该是围绕服务去开展应用,基于别人的平台,用别人的软件,使用外部的资源提供读者服务。无论是开源软件还是云计算技术,都应该强调服务这个核心思想,只有方向明确,工作才能更好开展。

### (2)具有国际视野

图书馆发展建设是一个动态的过程,它会随着技术、社会、经济等因素的发展而发展,图书馆的决策者在构架图书馆系统时,应该长远考虑,多学习借鉴国际图书馆的经验做法,预留扩展接口,为未来的发展提供充分的扩展性。新技术的发展非常迅猛,新技术思想层出不穷,需要及时准确预判技术发展趋势,保持同国际主流接轨。

### 参考文献:

- [1] IAN ROWLANDS AND DAVID BAWDEN. Digital Libraries: A Conceptual Framework [EB/OL]. [2008-02-18]. <http://www.librijournal.org/pdf/1999-4pp192-202.pdf>.
- [2] [5] Can Türker. Peer-to-Peer 数字图书馆体系结构中的对等、网格与服务定向[M]. 北京: 北京燕山出版社, 2005: 21-23.
- [3] Michael Stephens. Ten Trends & Technologies for 2009 [EB/OL]. [2009-04-18]. <http://tametheweb.com/2009/01/12/ten-trends-technologies-for-2009/>.
- [4] 奉国和. 多角度探讨数字图书馆发展演化[J]. 四川图书馆学报, 2008 (6): 21-25.
- [6] Ingo Frommholz, Predrag Knežević, Bhaskar Mehta, et al. Supporting Information Access in Next Generation Digital Library Architectures[C]. Preproceedings of the Sixth Thematic Workshop of the EU Network of Excellence DELOS, 2004: 49-60.
- [7] Walkerdine J., Melville J., Sommerville J.. A Framework for P2P Application Development, Technical Report COMP-004-2004[D]. Computing Department, Lancaster University, 2004.
- [8] 吴颖红. P2P技术应用与数字图书馆网络发展前瞻[J]. 晋图学刊, 2007 (3): 38-41.
- [9] 金海. 网格计算[A]. 中国计算机学会学术工作委员会. 中国计算机科学技术发展报告2004[R]. 北京: 清华大

- 学出版社 2004 99-119.
- [10]李永先.云计算技术在图书馆中的应用探讨[J].江西图书馆学刊 2009 39(1):105-106.
- [11]Michael Stephens. Ten Trends&Technologies for 2009. [EB/OL]. [2009-04-18].<http://tametheweb.com/2009/01/12/ten-trends-technologies-for-2009/>.
- [12]奉国和等.开放源码在数字图书馆建设中的应用[J].中国信息导报 2007 (4) 53-54.
- [13]奉国和.开源软件与图书馆知识管理探讨[J].图书馆论坛 2008 (2) 58-61.
- [14]奉国和.开源软件与图书馆创新能力研究[J].图书馆学研究 2008 (6).
- [15][26]钱国富,涂颖哲.基于开源软件的图书馆2.0建设研究[J].图书情报工作 2007 (5).
- [16]Jiawei Han, Micheline, Fan Ming, Meng Xiaofeng (Translation).Data Mining Concepts and Techniques [M].Beijing :China Machine Press 2001.
- [17]奉国和.基于SAS数据挖掘技术的航空CRM系统分析[J].情报杂志 2006 25(5) 56-59.
- [18]什么是SaaS?SaaS与传统软件区别是什么? [EB/OL]. [2009-04-14].[http://www.226e.net/article/91/Article40788\\_1.htm](http://www.226e.net/article/91/Article40788_1.htm).
- [19]Albanese A R. Campus Library2.0[J]. Library Journal , 2004 ,129(7).
- [20]Casey M. LibraryCrunch[EB/OL].[2008-03-16]. <http://www.Library.crunch.com>.
- [21]Boog J.Library 2.0 movement sees benefit s in collaboration with patrons[EB/OL].[2008-03-16].<http://www.publish.com/article2/0,1895,1881893,00.asp>.
- [22]孙茜.Web2.0对数字图书馆创新的启示[J].图书馆杂志 2005 (12) 27-30.
- [23][25]范并思,胡小菁.图书馆2.0 构建新的图书馆服务[J].大学图书馆学报 2006 (1) 2-7.
- [24]刘炜,葛秋妍.从Web2.0到图书馆2.0:服务因用户而变[J].现代图书情报技术 2006 (9) 8-12 67.
- [27]高岚岚.云计算与网格计算的深入比较研究[J].海峡科学 2009 (2) 56-57.
- [28]奉国和.网环环境下信息资源共享技术研究[J].图书与情报 2008 (4) 48-53.
- [29]霍小军.SaaS模式在电子政务中的应用初探[J].电子政务 2008 (1) 73-86.
- 作者简介:奉国和(1971-),男,博士,华南师范大学经济管理学院副教授,研究方向:数字图书馆,数据挖掘。

(上接第10页) [www.asp?sID=SpEXili&sYY=2009&ID=0BA6E367F0D6EJKJ8DGUJ4D6](http://www.asp?sID=SpEXili&sYY=2009&ID=0BA6E367F0D6EJKJ8DGUJ4D6).

- [2]Aims and objectives of proposed research[EB/OL].[2009-12-20].<http://booktrustadmin.kentlyons.com/downloads/CatherineHunn.doc>.
- [3]2008 Kids & Family Reading Report [EB/OL].[2009-12-20].<http://www.scholastic.com/aboutscholastic/news/readingreport.htm>.
- [4][7]Bookstart annual report[EB/OL].[2009-12-20].[www.nottinghamshire.gov.uk/bookstartannualreport0607.pdf](http://www.nottinghamshire.gov.uk/bookstartannualreport0607.pdf).
- [5]李超平.公共图书馆的阅读促进活动:重点目标人群与实施策略[J].公共图书馆 2009 (3).
- [6][8]吉姆·崔利斯.朗读手册[M].海口:南海出版公司, 2009 20-21.
- [9]Aims and objectives of proposed research[EB/OL].[2009

-12-20].<http://booktrustadmin.kentlyons.com/downloads/CatherineHunn.doc>.

- [10]陈永娴.阅读从娃娃抓起——英国“阅读起跑线”计划[J].图书馆理论与实践 2008 (1).
- [11]专家认为九年级以前做分级阅读标准非常重要[EB/OL].[2009-12-20].[http://edu.cyol.com/content/2009-08/11/content\\_2798231.htm](http://edu.cyol.com/content/2009-08/11/content_2798231.htm).
- [12]“中国儿童分级阅读参考书目”首批推出200种[EB/OL].[2009-12-20].[http://news.xinhuanet.com/book/2009-07/28/content\\_11785107.htm](http://news.xinhuanet.com/book/2009-07/28/content_11785107.htm).
- [13]袁晓峰等.快乐阅读进行时[M].贵州:贵州人民出版社 2009.

作者简介:李超平,女,浙江大学信息资源管理系教授。