

2005-2009年国内外数字图书馆 评价研究实践回顾与展望

□ 吴建华 / 华中师范大学信息管理系 武汉 430079

何汶 / 武汉大学图书馆 武汉 430072

摘要: 回顾2005—2009年国内外数字图书馆评价研究与实践活动, 介绍其中具有代表性的理论、模型、标准、系统、方法、案例, 指出国内存在的主要问题和对策, 总结数字图书馆评价的发展趋势。

关键词: 数字图书馆, 评价, 实践, 综述

DOI: 10.3772/j.issn.1673-2286.2010.03.005

1 国外研究与实践现状

由于数字图书馆本身的复杂性, 数字图书馆评价研究和实践的范围非常广泛, 参与这项活动的人员也比较多。从组织形式上看, 有机构组织的专题研究和实践活动、项目评价, 也有个人开展的一般理论与方法研究、实证研究。

1.1 机构活动与标准规范

一些行业协会和组织对数字图书馆评价问题高度关注, 组织各种研究活动, 提出评价标准, 建立评价工具, 并广泛开展评价活动对标准和工具进行检验, 使评价工具日趋完善, 并产生广泛的影响。主要有DigiQUAL、COUNTER、E-Metrics、eVALUED、SCONUL、EQUINOX、DELOS等, 这里仅介绍近5年有新进展的2个标准规范。

(1) DigiQUAL

美国研究图书馆协会在LIBQUAL+的基础上研究开发了DigiQUAL, 用于评价国家科学数字图书馆(the National Science, Math, Engineering and Technology Education Digital Library, NSDL) 为其用户群提供的服

务。DigiQUAL围绕数字图书馆服务质量相关的12个主题提出了180多个评价单元。经过2005年春夏的测试与分析, 2006年之后开始投入评估与试用^{①②③}。

项目组成员Kyrillidou在JCIDL2005数字图书馆联合大会和Northumbria2007图书馆信息服务绩效测度两个国际会议上介绍了开发DigiQUAL的起因、目标、方法, 以及对德克萨斯大学奥斯汀分校数字图书馆项目UTOPIA所进行的第一个实际评价的实施情况^[1,2]。

DigiQUAL™使用矩阵取样程序为每一个给定的图书馆生成一个汇总概要文件。为了减少被调查对象的工作量, 在DigiQUAL™内容丰富的题库中, 每位用户会面对随机产生的5个问题。每一位参与调查的人只需要完成随机选取的5个问题, 而不是必须回答每一个问题。由于问题是随机分配的, 很多人参加调查的结果是, 所有问题的响应比较均匀, 就好象每个用户回答了所有的问题。项目组2005年运用它对包括DLESE和MERLOT在内的NSDL的五个成员数字图书馆进行了试验评估。

德克萨斯大学奥斯汀分校开发的UTOPIA数字图书馆, 是第一个运用DigiQual™工具进行评价的数字图书馆。UTOPIA是一个把大学的知识、研究和信息向广大公众开放的一个数字图书馆, 资源丰富, 拥有极高

^① [2010-02-08]. <http://www.arl.org/libqual>.

^② [2010-02-08]. <http://www.libqual.org>.

^③ [2010-02-08]. <http://www.digiqua.org>.

的访问量,是最繁忙的英文版教育网站之一。UTOPIA的用户来自世界各地,如何让他们参与调查是一个问题。最初,UTOPIA开发团队在网站上建立了一个链接,让访问者回答网络问卷调查。但这种办法效果不好,于是借鉴网站开发阶段开展可用性测试的做法,让访问者填写个人信息,以发送礼品等形式激励访问者参与。UTOPIA把DigiQUALTM看作是一个持续的评估工具,收集信息、建立社区。尽管依然存在技术、方法及其他具体问题有待解决,但无论是UTOPIA把评估作为建立社区的手段,还是DigiQUALTM的实施,都取得了可喜的进步。

(2) COUNTER

COUNTER规范为图书馆等数字资源使用部门获取科学可靠的使用数据提供了根本保障。目前有两个版本,版本1是电子书和参考资料规范,2006年3月开始生效。版本2是电子期刊和数据库规范,2005年4月发布,2006年1月生效。版本3于2008年8月发布,取代版本2,要求最迟于2009年8月31日前执行新的版本^[4]。

版本3有4个变化。第一,通过采用新的协议,减轻联邦搜索引擎、自动搜索引擎、因特网机器人、网络搜索器等对使用统计的影响。第二,采用标准化使用统计收割协议(Standardized Usage Statistics Harvesting Initiative, SUSHI),促进对COUNTER使用统计的强化、管理和分析。第三,改善图书馆联盟使用报告。第四,改进期刊存档使用报告。

1.2 评价研究

(1) 一般性评价问题

萨拉赛维奇(Saracevic)教授一如既往地关注数字图书馆的评价问题,在2005年美国罗杰斯大学和克罗地亚奥西耶克大学共同主办的“数字时代的图书馆”国际研讨会上,他又一次提出数字图书馆的评价问题,并对80个评价研究从结构、环境、标准、方法四个方面进行总结^[3,4]。

美国雪城大学的Scott Nicholson提出整合图书馆各种资源和服务利用情况的数据建立数字图书馆使用数据仓储,开展书目挖掘活动,一方面为评价数字图书馆提供数据基础,另一方面为实际工作者的决策和研究人员深入研究数字图书馆利用模式提供数据基础^[5]。

美国威斯康星大学密尔沃基分校的Hong Xie研究

用户视角的数字图书馆评价标准和用户对数字图书馆的评价。他从该校信息研究学院招募了48位对数字图书馆使用感兴趣的本科生和研究生,在对他们进行必要的培训之后,由他们提出数字图书馆的评价标准,并运用自己提出的标准对一些实际运行的数字图书馆进行评价。数据分析结果显示,用户认为最重要的评价指标分别是:可用性、藏品质量、服务质量、系统性能、用户意见征集。用户视角的数字图书馆评价标准与研究人员提出的标准有明显的差异。用户用他们自己提出的标准考察5种类型的数字图书馆,发现了目前数字图书馆设计与开发方面存在的一些问题。调查还发现了一些对设计数字图书馆非常有帮助的意见,例如,由于数字图书馆的目标用户是不同层次的用户群,易于使用和用户控制同样重要。数字图书馆不仅要提供高质量的藏品,还要帮助用户判断藏品质量。此外,提供传统的和独特的服务对数字图书馆而言都是很重要的^[6]。

他让用户进一步使用并评价前期评价过的美国记忆(AM)和威斯康星大学数字收藏(UWDCs)两个数字图书馆,运用日记、调查问卷、调查的方法采集评价数据。评价过程分为3个步骤。首先,要求每位用户完成6个任务。然后,以封闭式问卷调查的形式要求用户对数字图书馆评价指标按自己认为的重要程度进行排序。最后,由用户运用他们认为重要的评价指标评价两个数字图书馆的质量,并排序。让作者感到惊奇的是,此轮研究发现,可用性和系统性能被认为是最重要的评价指标,而不是前一项研究中的可用性和藏品质量,因为用户充分相信此次评价的两个数字图书馆的藏品质量^[7]。

伊朗伊斯兰阿萨德大学的Alireza Isfandyari-Moghaddam和Behrooz Bayat运用文献调研的方法对数字图书馆的评价、管理与教育问题进行研究,发现数字图书馆实证评价研究集中在可用性、功能、可访问性、质量四个方面^[8]。

佛罗里达州立大学的John Carlo Bertot等连续三年反复对佛罗里达电子图书馆及其他数字图书馆的功能、可用性、可访问性进行评价,持续改进其对不同用户群的服务。这种多途径、反复的评价也更为丰富和健全,适于满足多样化用户群的需求^[9]。

英国斯特拉斯克莱德大学的Sudatta Chowdhury等着重对数字图书馆可用性、数字图书馆影响方面的文

^[4] [2010-02-28]. http://www.projectcounter.org/code_practice.html.

处理层由检索模块、抽取模块、计算模块三个模块组成。检索模块通过Web或本地文件系统获取必要的评价信息,收集记录有数字图书馆服务行为、元数据、数字对象信息的日志文件。抽取模块运用内置的或外部的解析器从收集的文件中抽取数据,并转化为5S质量模型的标准格式。解析器包括内容解析器(解析PDF等文件)、特定元数据格式解析器(解析Dublin Core等格式),以及特定日志格式解析器(解析XMLLog等格式)。界面层有配置模块负责存储数量指标,根据用户的选择生成XML报告和图表。他们运用自动评价工具对弗吉尼亚理工学院博硕士学位论文数字图书馆VT-ETD、巴西计算数字图书馆BDBComp、ACM三个实际运行的数字图书馆进行了评价,为每个评价对象的每一个属性生成直观的评价图表。他们邀请了3位可用性专家对自动评价工具的界面进行启发式评价,发现了53个问题;还运用邮件访谈和问卷调查的方法收集自动评价工具潜在用户对系统的评价意见和期望^[13]。

②数字图书馆成功模型

同样也是美国弗吉尼亚理工学院计算机科学系福克斯(Fox)领导的数字图书馆研究实验室的博士生沈(Shen),对贡柯维斯的数字图书馆质量模型进行了评判性的继承和发展,以数字图书馆质量模型、信息系统成功和接纳模型、信息搜索行为模型为基础,提出了数字图书馆成功模型。沈从最终用户的角度评价数字图书馆。他的基本思路是,当一个系统的预期用户在需要时经常性地使用系统,系统就成功了。用户满意会促进用户对系统的接纳,这会导致更高的系统利用率。因此,用户对数字图书馆的满意会导致数字图书馆的成功。

数字图书馆成功模型从四个方面评价数字图书馆:信息质量(IQ)、系统质量(SQ)、绩效预期(PE)、社会影响(SI)。数字图书馆是一个信息系统,由信息和系统两大部分组成,信息是系统的管理对象,系统是信息的管理工具,用户通过系统实现对信息的检索和利用。因此,信息和系统具有一定的独立性,信息质量和系统质量都会对数字图书馆成功产生重要影响,信息质量也会在一定程度上影响系统质量。绩效预期是指用户相信一个特定数字图书馆在帮助完成预期目标方面受益的程度。数字图书馆的社会影响是人们感觉到的其他重要人物对某数字图书馆的喜爱程度。最终让人们接纳并喜爱一个数字图书馆,主要取决于其信息质量和系统质量^[14]。

③费尔的数字图书馆评价新框架

德国杜伊斯堡—爱森大学的费尔教授(Fuhr)等对数字图书馆评价研究领域进行了全面考察,提出了数字图书馆评价的新框架^[15]。该框架认为任何数字图书馆都由系统、内容、用户组成,其中任何两个部分相互作用形成一个评价轴,共三个评价轴:性能(系统-内容)、可用性(系统-用户)、有用性(内容-用户)。

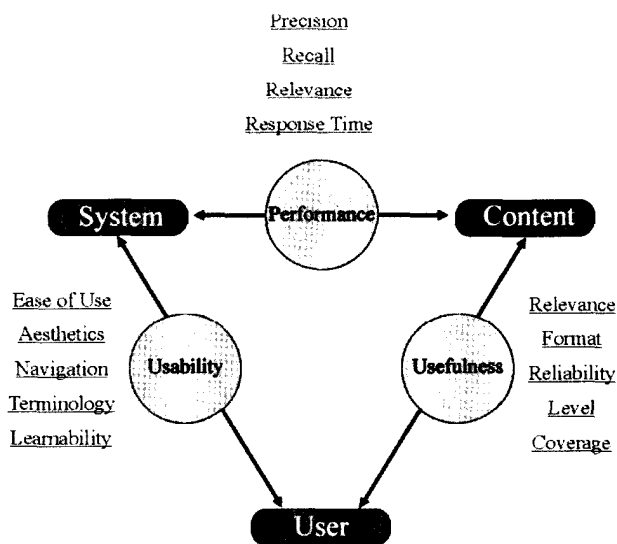


图2 费尔的数字图书馆评价交互三联框架
(Interaction triptych framework, ITF)

1.3 评价实践

(1) 对系统功能和性能的评价

①数字图书馆在不同检索模式下性能的比较评价

以色列本古里安大学的Robert Moskovitch等对一个半结构化临床指南文档数字图书馆进行全文检索、基于概念的检索和环境敏感检索的比较评价,研究不同检索模式对查准率的影响。结果表明,使用基于概念的检索和环境敏感的检索对提高查准率明显有帮助。一般情况下,查询语句中使用的本体元素越多,查准率越高^[16]。他们还设计并实施了一个通用搜索引擎Vaidurya作为临床指导数字图书馆DeGeL的组成部分。该数字图书馆除具备全文检索功能外,还可进行基于概念的检索和环境敏感的检索。这两种检索功能来自对临床指导文献的预索引和按照给定的本体进行半结构化处理。评价研究表明,两种检索是可行的且都能

够显著改善检索性能。评价还发现检索界面需要进一步优化,系统还需要支持基于概念的自动分类^[17]。

新西兰怀卡托大学的David Bainbridge等人评价音乐数字图书馆三种旋律检索模式的检索效率和有效性,在分析评价数据的基础上,对三种模式两两混合,发现混合算法产生更好的检索效果,即查全率和查准率都得到提高^[18]。

领域专家在特定领域数字图书馆中检索文献以满足特定目标的信息需求时,往往需要在很短的时间内找到有用的资料。波特兰州立大学的Susan L. Price等人开发了一个语义成分模型,在一个特定领域数字图书馆中加入部分带有语义成分的文献,研究在一个特定领域数字图书馆中加入语义成分是否会改善检索效果。研究者邀请30位领域专家分别针对4种预设情形检索所需文献,然后分别从系统中心角度和用户中心角度评价检索结果。从系统中心角度看,检索者使用含有语义成分的系统实施精确定位的检索时检索效果更好。从用户中心角度看,使用含有语义成分的系统时,检索者不得不输入更多的查询、花费更多的时间^[19]。

②数字图书馆辅助工具评价

Rodrygo L.T. Santos等人建立了一个新的用于建立组件化数字图书馆的基于Web服务的软件框架WS-ODL。这是一个三层框架:数据仓储-Web服务-客户端。数据仓储建立在Fedora上。在此框架中,传统的RDBMS技术可以很容易地用来支持几个通用数字图书馆服务。借助一个基于数字图书馆通用元模型的向导工具,对框架进行配置和定制,可以建造专门的系统。这个向导工具实现的工作流程帮助设计人员按照预定的步骤完成数字图书馆设计任务。对框架和向导工具的性能和可用性评价表明,框架可行且性能优良,向导工具在建立数字图书馆方面是有效的。评价实施了三个测试:安装测试、数据加载测试、性能测试。其中安装测试分为手工安装和使用向导工具安装。每种安装测试都有8人参加,其中4人是计算机专业的学生,4人是图书馆与信息科学专业的学生。执行安装要完成5项任务:A-安装并配置Fedora, B-安装Apache Axis, C-在Axis上安装Web服务, D-安装程序组件servlets, E-向结构化数字图书馆插入数据。安装测试表明:向导工具可以非常有效地帮助配置数字图书馆^[20]。

③数字图书馆特殊功能评价

宾夕法尼亚州立大学的Umer Farooq等设计并评价计算机与信息科学领域的学术数字图书馆CiteSeer的信息知晓(awareness)机制。评价分三步进行。第一,需求启发。在CiteSeer网站上征集参与者,通过网络调查和email访谈,发现CiteSeer的用户希望把RSS feeds作为通报系统以了解最新的出版信息如论文被引用情况、自己的研究领域新发表的论文、与自己论文相关的论文等。第二,原型评价。对三种RSS源即引用、新论文、相关论文的内容设计可行性进行形式化评价。这一阶段的研究发现了用户分别喜欢三种RSS源中什么样的内容。第三,自然研究。成立三个研究小组开展协作研究,他们在一个被称为BRIDGE的合作环境里共同针对计算机专业新入门学生编程教育问题写作新的见解。BRIDGE中有几个工具:用于交流的聊天工具、用来产生构想的头脑风暴空间、用来把构想可视化的概念地图、用于形成最终成果的意见件(opinion piece),其中头脑风暴空间和意见件均基于wiki。试验合作期间,每人先拿到一篇论文,他们以此论文为基础进行扩展、批判、形成自己的意见稿。然后,系统通过RSS分别推送给每个人新论文、引用第一篇论文的论文、相关论文。参与者通过文字聊天交流意见。任务完成以后,对参与者进行了半结构化访谈。研究发现,CiteSeer用户把RSS源集成到自己的合作环境中,让自己及时了解研究领域发生的情况^[21]。

美国伊利诺伊大学香槟分校的Michael B. Twidale等探索把数字图书馆的检索功能紧密集成到用户的学术写作环境中,把用户正在写作的内容作为生成检索词的来源,去检索可能会有用的内容。他们开发的工具就是PIRA。他们邀请了15位图书馆与信息科学专业的博士生试用PIRA。测试者与系统的交互以及他们写作的文稿都被记录下来做进一步分析之用。日志文件也被用于用户研究。初步的用户研究发现传统的评价方法并不适用。他们希望通过快捷的评价推动快速开发,结果是,评价活动过于复杂,绝大多数评价技术都过于缓慢^[22]。

个性化可以通过适应和自适应两种技术实现。英国布鲁内尔大学的Enrique Frias-Martinez等开发了一个满足不同认知方式用户需求的个性化数字图书馆,研究用户对两种不同技术的反应。个性化数字图书馆有两个版本,一个是可适应的(Adaptable DL),一个是自适应的(adaptive DL)。前者是用户通过调整策略适应系统,后者是系统通过观察用户的行为适应用户。

评价活动先通过问卷调查的方法了解用户的认知方式,使用的是专用问卷QUIS(用户界面满意度调查问卷)和CSUQ(计算机可用性调查问卷)。然后,让60位测试对象分别在两个版本的个性化数字图书馆中完成8个任务。数据分析表明,用户不仅在自适应系统中表现更佳,而且对自适应系统的感知更加积极主动。此外,认知方式极大地影响用户对适应和自适应的响应^[23]。

数字图书馆、电子期刊、电子文档等给研究人员带来便利的同时,也带来了困扰:他们没有时间跟踪、阅读他们能够获取的越来越多的期刊、会议论文、报告。这引发了一个新的需求:希望有工具帮助研究人员从这么多论文中建构意义。学术本体项目(Scholarly Ontologies Project)是英国工程和自然科学研究委员会(EPSRC)资助的项目,目的是开发一个论点服务器(claims server),支持学术争论和解答,让研究人员能够描述、争论某一篇文献的关键贡献,以及与其他文献的关系。该项目提供三个工具支持学术争论:ClaiMapper,让用户绘制单篇论文及论文之间的关系的观点地图;ClaiMaker,保存这些地图的数字图书馆服务器,提供解释服务以帮助查询多篇论文的观点地图;以及ClaimFinder,一个供新手使用的在ClaiMaker中进行检索的界面。

英国开放大学的Victoria Uren等人设计、实施并评价理解研究文献的意义建构工具。

他们开展了一个两阶段的用户评价活动,对一个多学科领域的文献进行分析,建立一个“论点网络”(claim network)。第一阶段建立“论点网络”(claim network)模型,写一个关于这个论题的简短综述。第二阶段是对传统的文献综述和“论点网络”(claim network)进行比较评价。两组学生参与评价。一组利用传统的文献分析工具即手写的文献综述,另一组就直接使用这个“论点网络”(claim network)。他们解答与该领域有关的提问。结果表明,“论点网络”(claim network)可以支持回答诸如“这个观点来自哪里?”、“什么证据支持这个观点?”之类的简单问题。研究还发现,“论点网络”(claim network)帮助测试对象形成自己的观点。总之,评价发现,“论点网络”(claim network)可以帮助研究人员进行文献分析^[24]。

④数字图书馆开源软件综合评价

新加坡南洋理工大学的Dion Hoe-Lian Goh等,根

据建立亚洲通讯资源中心文献数字图书馆的实际需要,在文献调研和需求分析的基础上,运用德尔菲法确定评价指标和权重系数,建立了包括12个一级指标的三层开源数字图书馆软件综合评价指标体系。12项一级指标评价是:内容管理、内容获取、元数据、搜索支持、访问控制与安全、报告与查询功能、保存、互操作性、用户界面、遵循标准、自动化工具、支持与维护。运用该评价工具对CDSware、EPrints、Fedora、Greenstone四个流行开源软件进行评价,发现Greenstone最符合12项评价标准,也最能满足作者开发数字图书馆的需要,CDSware的评价总分与Greenstone十分接近,只是在安装平台和花费的安装时间方面稍稍逊色。这两款软件都有5项一级指标得满分。Fedora有4项一级指标得满分,特别是在数据保存和支持标准方面优势突出,但在自动工具和内容管理功能方面较弱。EPrints在内容获取和内容格式支持方面具有优势,但缺乏报告和查询功能,特别是搜索功能较弱,只支持元数据检索,不支持全文检索。四款软件在用户界面和对标准的支持两个一级指标上都得了满分。该评价指标体系可以根据需要加以调整,用于其他需要评价数字图书馆软件的情形。这种严格、规范的评价实践非常值得借鉴,建立的开源软件评价指标体系也可加以利用^[25]。

(2) 可用性评价

①网站可用性评价

萨拉赛维奇教授指导的博士生朱迪·金(Judy H. Jeng)建立了一个数字图书馆可用性评价模型,并以两个大学图书馆的网站为对象进行了实证研究。数据收集阶段使用了形式化可用性测试技术,运用了问卷调查、执行任务、观察、有声思考等方法。设计了信息搜寻任务,把试验对象的操作过程记录下来并对记录进行分析,通过计算时间、步骤等实现量化评价。同时,借助测评数据,她还对有效性、效率、满意度之间的关系进行了深入研究^[26]。

科罗拉多州立大学的Don Zimmerman和Dawn Bastian Paschal对该大学数字图书馆和西部水域数字图书馆进行了可用性评价,18位测试对象访问两个数字图书馆网站,在每个网站上完成11项任务。测试对象对两个数字图书馆的评价是积极的,但分别有25%和36%的测试对象不能完成所有任务。评价发现的问题被用于改进网站设计^[27]。

柏林洪堡大学的Najko Jahn从人类学动机角度评

价新开发的国际关系与地区研究数字图书馆IREON的可用性。在可用性评价的不同阶段,针对不同的用户群,分别运用焦点小组、启发式评价、用户测试方法收集评价信息。焦点小组方法针对国际关系与地区研究专业的学生,启发式评价收集信息专家(信息研究、信息服务、网站管理)评价数据,用户测试则针对政治学学科专家和研究生。通过组合运用多种方法,从不同层面的用户(本科生、研究生、学科专家)和信息专家(信息研究人员、信息服务人员、网站开发人员)的角度发现了很多可用性问题,显著改进了数字图书馆IREON的可用性。这种多方法并用的可用性评价实践被证明是可行且成本低廉的,对现在通用的可用性评价方法是一个有益的补充。研究还发现,信息专家和网站设计人员关注不同的可用性问题,而政治学的研究人员和学生发现的可用性问题则具有高度的一致性^[28]。

②系统原型可用性评价

随着越来越多的大学建立了数字机构库,机构库成为重要的信息资源。韩国教育与研究信息服务部门KERIS为了让用户方便地利用这些资源,自2003年起建立韩国大学机构库联盟dCollection,同时开发了机构库软件平台,免费提供给62个成员大学使用。明知大学的Hyun Hee Kim和金庆国立大学的Yong Ho Kim在文献调研的基础上提出一个包括满意度、支持性、有用性、效用四个一级指标的机构库可用性评价标准。运用该标准对30个测试对象(24位研究生和6位教授)进行测试,开展了两个可用性实验室和远程测试实验,对4位专家进行了一次焦点访谈。每位测试对象要完成4项规定的任务,他们之前都没有使用过联盟的系统。运用评价结果提出改进联盟系统可用性的具体建议,包括主页、元数据提交、检索与浏览几个方面。研究发现,针对如何注册、如何检索文档提供有效的指导,改善视觉效果,基于FRBR模型聚集并显示相关资料,对于提高系统的可用性具有重要作用^[29]。

美国科罗拉多大学的Holley Long等对科罗拉多历史航拍照片库原型进行可用性评价,该评价测试的结果用于对原型进行“以用户为中心的再设计”。评价经过了三个阶段:需求分析、启发式评价、用户测试。需求分析阶段,首先分析照片库的目标用户,然后对部分目标用户进行访谈,在此基础上,提出系统的目标、局限性和功能需求。所有的访谈对象都要求系统提供浏览、下载不同分辨率照片的功能,这显然

是一项重要功能。启发式评价阶段,招募了6位评价人员,利用尼尔逊(Nielsen)提出的广为接受的10个可用性启发问题,以及施乐公司(Xerox)的“启发式评价-系统清单”,分别去发现原型存在的问题。结果,评价人员共提出了83条意见,其中43条内容是重复的。用户测试阶段,有2名教师和3名学生志愿参加了测试,每位测试对象完成5项任务。评价发现了功能、导航和布局等方面的问题,被列为下一步的设计目标。评价活动还发现了研究人员是如何使用航拍照片的,以及他们偏好的访问方法、期望的功能,这对于改善系统、更好地服务用户无疑具有积极意义^[30]。

③数据库可用性评价

印第安纳大学的Giannis Tsakonas等评价一个开放存取数字图书馆E-LIS的有用性和可用性。他们采用费尔(Fuhr,N.)提出的交互三联框架ITF开展评价,研究者通过在线问卷调查收集数据,问卷包括34个问题,共获得131个有效问卷^[31]。

新西兰罗杰斯大学的Xiangmin Zhang等对ACM、IEEE CS、IEEE Xplore三个计算机领域的数字图书馆的检索与浏览特征进行可用性测试研究,12名计算机科学与工程专业专业的本科生、12名计算机科学与工程专业的研究生、11名图书馆与信息科学专业的研究生志愿者参与了测试。研究发现,不同的交互设计对用户使用数字图书馆的效果有显著影响,同时还发现了一些易导致不良用户使用效果的交互设计,可用于改进数字图书馆的交互设计^[32]。

④可用性与有用性组合评价

英国斯特拉斯克莱德大学的Steven Buchanan和Adeola Salako认为系统的可用性和有用性是系统反映交互性的两种相互依存的属性,前者是关于用户界面,后者是关于支持用户活动的功能,二者共同决定了系统的满意度和使用情况。但现实中人们往往只关注数字图书馆的可用性。为此,他们在文献调研的基础上,提出了一个测度数字图书馆可用性和有用性的组合模型,并对一个健康服务数字图书馆进行了试验评价。模型中体现可用性的关键属性有:有效性、效率、美感度、术语、导航、可学习性;体现有用性的关键属性有:相关性、可靠性、时效性^[33]。

⑤少儿数字图书馆及跨语言跨文化评价

国际儿童数字图书馆(ICDL)是一个跨文化、跨语言的、以3-13岁儿童文学作品为内容的数字图书馆项目,由美国马里兰大学人机交互实验室实施。ICDL社

区项目是ICDL项目的扩展,目的是以ICDL的馆藏图书为基础,开展基于课堂的儿童社区活动。具体地说,是让7-11岁的来自不同文化背景、说不同语言的儿童,不借助机器翻译,就能开展与ICDL的图书相关的活动。参与活动的儿童阅读ICDL中本国语言的故事,通过绘图制作个人基本信息,进而与其他国家的儿童交流。他们还可以创作故事,并与其他儿童交换,就故事提问或解答他人的提问。系统通过基于图形的用户界面工具和语词翻译支持这些活动。系统原型提供的功能有创作故事、交换故事、相互认识。研究者组织开展了三轮评价活动:美国-匈牙利、美国-阿根廷、美国-墨西哥,参与者的任务是制作介绍、读故事、为故事创作新的结局、交换故事、解释故事、提问和答问等。研究人员在一旁观察、笔记,对部分过程录像。通过这种评价活动,发现了很多可用性问题的,为改进系统界面提供了依据^[34]。

美国马里兰大学哈切森等人针对现有数字图书馆主要为成人设计,基本不考虑少儿的特定习惯和需要的实际情况,以国际少儿数字图书馆(International Children's Digital Library, ICDL)为研究对象,选择6—11岁的少儿为测评对象,研究数字图书馆可用性问题^[35]。针对一般用户的数字图书馆,基本不考虑少儿信息处理技能、搜索和浏览技能,以及少儿特定的信息搜索偏好,因而设计出的数字图书馆,并不一定适合少儿用户。研究者认为,一般数字图书馆中广泛使用的层次式分类浏览界面对少儿来说过于抽象,使用起来比较困难,平面分类结构更适合这个年龄段的用户。但是,这样一来,界面上需要容纳的内容就太多,需要设置翻页、滚动功能,对可用性问题造成重要影响。为权衡利弊,他们分别设计了层次式浏览界面和平面式浏览界面,然后设计可用性任务进行测试和对比研究,发现平面式界面无论是用于布尔检索还是随意浏览,都更加优越。这一结果对设计适合少儿的数字图书馆界面具有指导意义。设计界面时,他们充分考虑了少儿的特定需求和习惯。例如,少儿操作鼠标和移动光标比较慢,定位也比较困难,故界面上的图标要大一些;检索与浏览中常用的关键词和层次分类目录都不适合少儿,需要用图像等替代;少儿选择图书的标准和方法都不同于成人,他们更多地从外观、图像、特定类别的动物或游戏等角度加以选择。设计时,少儿用户的这些特性都要考虑进去。

⑥用户与系统契合评价

伦敦大学学院的Ann Blandford等研究以用户为中心的图书馆评价,特别是关注数字图书馆与用户的信息工作环境的契合,关注数字图书馆如何支持用户的信息工作。他们运用PRET A Reporter框架构建以用户为中心的评价研究,并对3个数字图书馆进行了可用性评价。运用PRET A Reporter框架开展评价研究要按以下6个步骤进行:第一步,明确评价目的(Purpose of evaluation);第二步,弄清可用的评价资源和局限性(Resources and constraints);第三步,考虑相关伦理问题(Ethics);第四步,确定数据采集技术(Techniques for gathering data);第五步,选择分析技术(Analysis techniques);第六步,报告评价结果(Reporting of findings)。

第一个评价是对Garnet——一个基于Greenstone的空间超文本数字图书馆的形成性评价,运用屏幕录像和日志文件采集数据。第二个评价是对一个供人文学者使用的数字图书馆的两个不同界面进行评价,考察数字图书馆设计如何支持用户更大范围的工作活动。记录20位测试对象的有声思考并录像。第三个评价是了解专业人士如何利用数字图书馆支持其日常研究与写作任务,通过访谈25位记者采集评价数据。3个实证评价证实该评价框架在规划和实施评价方面是十分有效的^[36]。

他们还运用一种新的可用性评价方法——基于概念的表面与结构错位分析(CASSM, Concept-based Analysis of Surface and Structural Misfits),系统地分析用户与系统之间的概念错位。与现有评价方法如启发式评价、认知演练等不同的是,CASSM重点关注概念结构而不是过程。该方法关注用户的概念模型与系统实施之间的契合质量,既有数据收集,又有数据分析。它借助一个开源工具Cassata进行CASSM分析。

用户与系统之间之所以存在概念错位,是因为用户对系统形成了不精确的心智模式(改进系统设计可以解决这一问题),或者系统的实施是基于不精确的对用户工作环境的表达。CASSM分析是为了发现这种错位,重新设计系统,从而让用户概念模型与系统的实施之间更好地契合。CASSM分析两类概念错位,一种是表面错位(Surface misfits),一种是结构错位(Structural misfits)。表面错位突出用户的概念,部分或完全地偏离系统概念,这样一来,用户在学习和使用系统时就会碰到困难。结构错位的情况是,用户概念模型与系统概念模型之间的关系不一致,其结果

是,系统的改变会造成用户的困难。

作者分别对机械臂、数字图书馆(ACM DL)、绘图工具三个交互系统进行评价,演示CASSM的不同方面和不同的分析深度。其中对数字图书馆ACM DL的评价邀请4位人机交互研究领域的研究生作为测试对象,研究人员运用有声思考法观察他们的检索行为和过程^[37]。

(3) 使用评价

①调查统计法

土耳其安卡拉大学Atilgan D.和Bayram O. G.运用问卷调查的方法对教师了解和使用数字图书馆的情况进行了评价^[38]。

由于问卷设计得非常简短,被调查教师只需要花费很少的时间即可完成问卷填写。他们对安卡拉大学所有3800名教师发放了问卷,要求三天内返回给图书馆。结果有2100名教师按要求返回问卷,其中1996份问卷可供统计分析之用。

对调查数据统计分析,发现了一些重要事实:多数教师对学校的数字图书馆有所了解,很多人使用数字图书馆;最受欢迎的三个数据库是:Web of Science、ScienceDirect和EBSCO;医学、科学、技术类专业的教师使用数据库最多;使用数据库的目的主要是研究(86.7%),其次是教学(11.9%)和一般性了解(1.4%);在图书馆网站上链接帮助信息(38.1%)是教师们最喜爱的帮助方式,其它形式分别是提供辅导材料、培训课、咨询馆员,等等。

作者从调查结果得出结论,认为图书馆需要改善培训和宣传服务,以增进教师对数字图书馆的了解。他们还对图书馆建立了数字图书馆之后学校教师发表论文的情况进行统计,数据显示数据库的使用对教师发表论文有显著影响,学校订购电子数据库之后两年,教师发表的论文数量显著增加。

为进行对比研究,他们又实施了一轮调查,发放了3500份调查问卷,返回的有效问卷为1024份。对比研究发现,了解数字图书馆的教师显著增加了,很多教师“经常”而不是“偶尔”使用数据库,副教授、助理教授、研究助理了解和使用数据库的程度最高。对比研究还表明,数据库培训在促进教师对数据库的了解和使用方面发挥了积极作用,教师们对联机网络培训更感兴趣。针对这一发现,学校决定由远程学习中心开发培训教程,然后放到网上供教师访问使用^[39]。

②日志分析与网站调查相结合的方法

Diane Harley和Jonathan Henke运用事务日志分析和网站调查两种方法研究数字图书馆的使用状况。他们的研究不仅发现了网站用户组成、行为模式以及网站资源被利用情况,而且对这两种方法相互配合使用进行了有益的探索,发现了单独运用一种方法存在的问题^[40]。

事务日志分析和网站调查各有优势和劣势,相互补充,可用于了解事物的不同方面。日志分析自动记录用户行为信息,不需要用户参与就可以了解各种资源的实际利用情况,以及用户的真实行为和使用模式。日志分析法的弊病是完全依赖客观记录的日志数据,但这些数据无法揭示一些重要信息,如用户的真实身份,基本情况如教育、专业、行业背景等。日志分析工具通过从日志文件中提取IP地址、客户域名、IP范围等确定用户身份。实际上,由于代理服务器的广泛使用、公共场合多人使用同一台计算机、同一用户分别从不同地点访问、缓存的影响等原因,导致这种确定方法不真实、不可靠,也不是用户的每一个行为都被记录下来了。调查法可用于了解用户的个人信息,以及态度、行为、动机,访问网站的真实原因,了解他们的满意水平,但存在响应率低,用户回答问题可能不真实,不具有代表性,对结论的形成产生偏见、误导等问题。

他们使用两种方法对加州伯克利大学建筑系的图片库SPIRO(Slide and Photograph Image Retrieval Online)和杰克·伦敦数字图书馆(The Jack London Collection)进行研究。具体方法是,把网络调查放到两个网站上两个月,这期间同步收集日志数据。调查问卷收集用户的身份、所属机构、访问这两个网站的频率等信息。对日志数据进行分析发现,大部分用户是第一次访问这两个网站,其中重复访问SPIRO的用户比例为11%,而杰克·伦敦数字图书馆的重复访问用户比例为17%。大部分用户通过搜索引擎到达这两个网站,比例分别为26%和31%,通过加州伯克利大学访问这两个网站的用户比例分别为30%和29%。通过统计回答问卷的数量和日志中唯一IP地址的数量,计算出网络调查的响应率,结果发现响应率都只有1%。这让研究者对调查对象的代表性和调查结果的价值产生了怀疑。

为了研究调查响应是否具有代表性,他们对三个行为测度指标进行了深入研究:日志记录期间每

人的浏览会话次数,每个会话访问文件的数量,平均会话时间长度。这三个指标对网络调查响应者和不响应者都可以计算出来。第一个指标测度对网站的使用频率,另两个指标测度用户对网站的使用深度或参与度。研究者比较两组用户在三个指标上的水平,从统计数据可以发现对网络调查响应的用户与不响应的用户相比,具有显著的差异:前者更频繁地访问网站,会话时间更长更深入,每个会话使用更多的文件。也就是说,对网络调查响应的用户不具有代表性。因此,不能以调查数据为基础得出有关网站访问者具有什么特征的结论。

(4) 其他方面的评价

①用户的评价

印第安纳大学的Elsa F. Kramer通过观察70个志愿者对一个摄影图像数字图书馆的使用,来测度和评价其功能、内容,以及他们对该数字图书馆的了解。在用户完成搜索任务之前和之后,通过简单提问的形式收集观察对象的基本信息和他们对数字图书馆的评价和建议。收集的用户数据还用于改进该数字图书馆的功能和元数据,选择合适的图片丰富藏品内容,以及制定市场化计划^[41]。

②评价数字图书馆对人的影响

格拉斯哥大学的Ian G. Anderson通过问卷调查、焦点小组、日志分析、网络调查等方法评价数字图书馆项目“格拉斯哥故事”对用户的影响。“格拉斯哥故事”(The Glasgow Story)是一个旨在促进终生学习的文化遗产数字图书馆项目,由位于格拉斯哥的多所大学、图书馆、档案馆、博物馆、教育管理部门等合作开发。评价跨越形成性评价和总结性评价两个阶段。形成性评价阶段,组织了多个研究领域的人员参加,主要对数字图书馆的内容、功能、网站等进行评价。总结性评价阶段主要评价该项目对人们的影响,从中学生到退休人员各个年龄段的人员都有参与。调查发现,青少年使用此数字图书馆的并不多,中老年人是其主要用户。网络调查因为反馈调查表数量太少,中途对调查表重新设计,情况有所改变,但数量依然有限。研究发现,此类评价应该在多个项目之间进行比较才更有意义,对单个项目的评价有其局限性。开展数字图书馆影响评价的实践并不多,此项评价也碰到不少实际困难^[42]。

③评价数字图书馆的服务质量

Cecilia Garibay等把质量功能展开系统(Quality

Function Deployment, QFD)与卡诺模型(Kano model)结合起来评价墨西哥瓜达拉哈拉大学数字图书馆的服务质量。通过网络问卷调查收集评价数据,建立了瓜达拉哈拉大学数字图书馆的QFD-Kano模型,清晰地呈现出用户感知的数字图书馆服务质量和用户的期望。该方法帮助辨明哪些需求应该优先满足,从而提高用户满意度。该方法还提供了满足用户技术需求的信息^[43]。

④评价虚拟参考是否遵循标准规范

美国印第安纳大学的Prina Shachaf和奥古斯坦纳学院的Sarah M. Horowitz评价虚拟参考咨询服务遵循行业标准的情况。他们采集54个图书馆的324个参考咨询实例,分析其是否遵循美国图书馆协会参考咨询与用户服务标准RUSA和IFLA的数字参考指南。结果发现,遵循标准的比例很低,而且用户满意度与是否遵循标准之间不存在相关关系^[44]。

⑤使用评价中日志的技术处理

评价网站使用情况通常需要计算一个会话期间浏览页面的数量。但由于根据日志文件无法确定会话何时结束,传统的做法是假定所有页面浏览都花费大致相同的时间。实际上,页面浏览时间差异很大,一篇论文可能要看2小时,一个菜单可能只需要不足10秒。这样一来,对会话的计量就非常不准确。伦敦大学学院的Paul Huntington等通过分析OhioLINK期刊数字图书馆的日志数据,说明超时定界符如何影响页面浏览时间的估算和会话数量的计量。

针对这一问题,他们提出根据浏览页面的类型设置不同的超时信号阈值。代理服务器会引发类似的问题,因为一个用户可能会接管上一个用户的会话。由于代理服务器的使用,一个IP可能对应一大群用户。为了解决会话计量不准确的问题,他们提出把500作为一个分界值,凡14个月内浏览页面不超过500个的IP,可能就不是代理服务器,就使用一个不那么严格的超时信号阈值。对代理服务器IP,就要使用较为严格的超时信号阈值^[45]。

⑤评价机构库的规模和使用的软件系统

印度科学通讯与信息资源国立研究所的Rekha Mittal和G. Mahesh运用文献调研和网络调查的方法调查印度公共领域的数字图书馆和机构库数量,以及其中藏品的规模,发现利用开源软件建立数字图书馆或机构库的情况非常普遍,最常用的几个软件分别是DSpace、EPrints、GreenStone DL,但是印度主要的数

评价标准,并对部分图书馆进行实证评估分析^[53]。

(2) 评价标准

评价标准或评价指标体系始终是一个研究重点。

刘炜等以城市图书馆的目标、职能为背景,以传统图书馆的评估方法和指标体系为参考坐标,提出复合型图书馆评价参考标准^[59]。

黄如花就数字图书馆信息组织提出从信息组织对象的采集、信息的揭示与描述、知识组织体系的建立、海量数据的集成、海量数据的搜索、信息的呈现、信息安全、用户界面八个方面构建评价指标体系^[60]。南开大学中国图书馆网站评价研究组对我国图书馆网站评价的现状、基本原则、评价指标体系、评价方法,以及公共图书馆和高校图书馆在评价研究方法的不同倾向进行了全面分析^[61]。肖珑等论述了CALIS数字资源评估指标体系的评估对象和评估内容,列出了完整的指标体系、指标属性、应用层面,并就应用方法给出指导意见^[62]。

王启云针对高校数字图书馆建设的现状,采用定性分析与定量分析相结合的方法,以传统图书馆的评估方法和指标体系作为参考坐标,提出一套适用于复合型图书馆进行数字图书馆建设和服务评估的综合指标体系^[63]。

李枫林介绍国外电子服务质量评价模型,根据综合度量模型和DigiQual+与数字图书馆服务质量相关的12个主题,提出一个数字图书馆服务质量评价模型,从环境质量、过程质量、结果质量三个方面来度量用户感知的数字图书馆服务质量^[64]。

张娴等介绍了国内外数字参考咨询服务质量评价体系的研究与应用概况,国家科学数字图书馆网络联合数字参考咨询服务质量评价指标体系和评价制度^[65]。

叶肖惠在梳理已有评价标准的基础上,提出以用户为中心的网络环境下参考咨询用户满意度评价指标体系,并提出可行的评价方法^[66]。

(3) 评价方法

对评价方法的研究和运用是数字图书馆评价研究的一个重要方面。

其中,对层次分析法的研究和运用较多。此外还有对数字图书馆知识产权各种评价方法及其适用性的分析^[67,68]、对国外数字图书馆绩效评估模型如SERQUAL、LibQUAL+AM在我国适用性的理论与实证研究^[69,70,71]。

吴建华从事物的信号显示机制出发,研究评价活动的信息原理,按照初始评价信息的来源构建数字图书馆评价方法体系。初始评价信息来源于人的主观判断的是主观评价法,具体有专家评价法和用户评价法。初始评价信息来源于客观事实的是客观评价法,即信息测度评价法。初始评价信息来源于两个方面的,是综合评价法^[72]。对每一种评价方法,都进行了系统深入的研究,并结合应用实例或实证研究说明其运用过程^[52]。

评价信息采集与处理方法是数字图书馆评价中的一个重要问题。数字图书馆服务器上存在大量有关用户访问信息的日志文件,但这些文件记录的信息内容与格式差异很大,不利于日志分析工具的互操作与结果重用。为了解决这个问题,吕娜提出了一个基于XML的数字图书馆日志标准格式,基于该格式的数据采集系统可以记录丰富的用户行为数据。她将该日志格式应用于实验系统,通过用户使用该实验系统收集一定量的用户行为数据^[73]。她还设计了基于频繁下载集的算法,对收集的用户日志文件进行试验性挖掘分析,得出用户的频繁下载集合,可用于推荐链接和评价数据库资源^[74]。

2.2 评价实践

(1) 资源评价

电子资源及其使用评价是我国数字图书馆评价的一个热点问题。

① 电子资源绩效评价

徐革对我国大学图书馆电子资源绩效评价方法进行了系统的理论和实证研究。她以服务质量差距理论GAP、SERVQUAL评价模型,和以用户感知为中心的图书馆绩效评价模型LIBQUAL+为理论基础,结合灰色理论、模糊理论和方法,使用专家群体调查法确定重要影响因子,建立了大学图书馆电子资源绩效评价模型和指标体系,提出利用网关设计来解决网络统计数据规范化采集问题。通过对我国211大学图书馆馆长进行专家咨询,调查电子资源绩效影响因素,构建了内容、检索系统、利用、成本、服务五个方面的电子资源绩效评价指标体系,并对Elsevier数据库、CNKI、超星数字图书馆进行了实际评价^[48]。

唐琼将电子资源使用评价划分为微观层面的特定类型电子资源使用评价与宏观层面的图书馆整体电子

资源使用评价,提出基于用户满意度的图书馆整体电子资源使用评价模型,利用该模型设计问卷对中山大学图书馆用户进行了满意度调查。运用相关分析与回归分析方法研究模型中4个自变量与用户价值感和用户满意度之间的相关性影响程度,以此为基础对模型进行修正,并运用象限分析法,为中山大学图书馆改进电子资源建设提供参考性建议^[75]。

②开放存取资源的评价

中南大学刘海霞、方平、胡德华等对开放存取期刊的质量进行了系统研究。他们利用DOAJ(Directory of Open Access Journal),从学科分布、影响因子、即时因子、引用影响,及即时因子与影响因子的排名比较,研究开放存取期刊的质量^[76]。对OA期刊与非OA期刊从论文的影响力、论文间质量的差异性、作者合作度三个方面进行对比研究,结果显示,OAJ论文已经达到甚至超过非OAJ论文的质量和影响力^[77]。分析同行评议法、引文分析法、网络计量学方法用于开放存取期刊质量评价的优势与不足,建议运用综合评价方法评价开放存取期刊质量^[78]。侯集体分析国内开放存取期刊质量评价研究现状,提出采用定性定量相结合的方法评价开放存取期刊质量,并给出具体的评价指标^[79]。

③电子资源采购决策支持评价

庞弘燊、徐文贤针对图书馆数字资源选购、续订、停订的决策需要,构建了一个由“数字资源购前评价与优选指标体系”和“数字资源购后评价优选指标体系”组成的“数字资源评价与优选指标体系模型”,并提出资源评价与优选的软件系统模型^[80]。

毛莉和陈惠兰利用统计分析本馆外文电子资源的质量和成本/效益,同时利用Scopus引文分析系统,对本校师生学术研究的论文引证情况进行统计分析,从而确定数据库供给与师生实际信息需求的差异,为数据库的续订或调整工作提供决策依据^[81]。

张彬以CALIS引进的外文期刊全文数据库为研究对象,以“CALIS引进数据库用户满意度调查总结报告”为依据,运用层次分析法构建外文电子期刊全文数据库评价指标体系,评价指标包括:基础指标、技术指标、效益指标。对3种外文期刊全文数据库Springer、Wiley、Blackwell进行了实际的测评分析,作为采购决策的依据^[82]。

(2)用户满意度评价

近几年我国出现了运用美国顾客满意度模型ACSI

对我国图书馆、数字图书馆进行评价的研究和实践活动。

吕娜、余锦凤构建了以用户为中心的用户满意度模型,对某大学数字图书馆进行了满意度测评,发现印刷型出版物、网站可用性、工作人员素质是影响用户满意度的主要因素,可以从这几个方面努力,改善总体用户满意度^[83]。

南京理工大学甘利人、李莉等对用户满意度测评问题进行了系统研究,构建了科技数据库网站顾客满意度指数模型ICSI,并进行了实证调查和分析,对模型应用流程进行了系统归纳总结^[84,85,86]。丁晟春等在基于用户满意的图书馆网站评价研究的基础上,设计了图书馆网站满意因素自主评价软件。利用该软件,各机构可根据自身需要确定评价方案,并根据系统提供的指标体系模板自动创建针对专家和普通用户的调查问卷,利用软件对调查数据进行统计分析,生成直观的三维图形,把网站诊断结果呈现出来,供改进网站之用^[87]。

(3)网站及可用性评价

黄晓斌在理论探讨的基础上对用户界面进行了实证评价研究^[88]。

向建军运用德尔菲法构建了由标识系统、组织系统、导航系统、检索系统4个一级指标组成的我国高校数字图书馆可用性评价指标体系,选取西南地区10所高校图书馆网站作为测试对象,使用问卷调查的方法获取评价数据,得到10所高校图书馆网站的可用性评价结果,并根据评价结果给出提高我国高校数字图书馆可用性的建议^[89]。

南开大学中国图书馆网站评价研究组分别针对高校图书馆和公共图书馆的特点,运用专家咨询法和层次分析法设计了高校图书馆和公共图书馆网站评价指标体系,并分别选择首批进入211的15所高校图书馆网站、29个省级公共图书馆网站作为评价对象进行了实证评价,分析这些图书馆网站的优点和存在的主要问题,并给出了相应的建议^[90,91]。

复旦大学图书馆赵宇翔在对公共图书馆网站可用性考察的基础上,提出了公共图书馆可用性评价指标体系,并对上海图书馆的网站进行了可用性测评。他还对复旦大学图书馆新版网站的可用性进行了形式化测评,设计了10个问题,选择了16位测试对象,开展了测试前问卷调查和测试后问卷调查。他们的可用性测试尽管条件有限,但比较规范^[92,93]。

(4) 数字参考咨询评价

过仕明在系统考察国内外研究成果和案例的基础上,利用层次分析法、专家咨询法和模糊数学方法,构建了数字参考咨询服务质量评价指标体系,对哈尔滨师范大学图书馆的数字咨询参考服务进行了测度和评价^[49]。

王晶晶在总结国内外研究实践的基础上,提出了一个由咨询服务系统、咨询服务过程、成本效益、咨询服务管理、用户满意度5个一级指标组成的数字参考咨询评价三层指标体系,运用专家调查法和层次分析法确定指标权重系数,并对某大学图书馆的数字参考咨询服务质量进行了实证评价,分析取得的成绩和存在的问题,提出提高数字参考咨询服务质量的建议对策^[94]。

虽然参考咨询评价方面的期刊论文比较多,但基本上是对国内外标准规范的介绍、对指标体系和评价方法的探讨,鲜有实证评价。

(5) 信息安全评价

赵坚在全面考察国内外信息安全风险评估问题、标准、软件研究的基础上,分析了实现基于ISO27000的数字图书馆风险评估软件的必要性和可行性,结合数字图书馆信息安全管理的特点,研究了基于ISO27000的数字图书馆信息安全风险评估辅助工具的流程,设计并实现了基于ISO7000的数字图书馆信息安全风险评估辅助工具。将数字图书馆的信息资产分为数据文档、硬件资产、软件资产、人员和服务五类。完成数字图书馆信息资产评估后,就应该对其威胁和薄弱点进行评估。在识别出数字图书馆中所有资产面临的全部威胁后,为每项资产面临的威胁赋值。系统将威胁权值分为1至5五个级别^[95]。

朱晓欢以信息安全管理标准系列ISO27000下的ISO27001标准和ISO17799标准提出的风险评估方法和信息安全管理措施为指导,对复合图书馆实施信息安全风险评估。归纳整理出复合图书馆信息资产-威胁-薄弱点对照表。对部分已实施数字图书馆项目的19所高校图书馆、公共图书馆、科学图书馆的信息安全状况调研,依据ISO27000的风险评估标准内容对调研对象进行风险分析和评估,以获得的数据为基础,建立复合图书馆信息安全的风险评估实施模型,根据ISO27000提出的风险评估步骤,结合复合图书馆的业务特点,分别从信息资产、威胁和薄弱点三个安全要素对复合图书馆的信息安全风险进行分析和评估,对

某高校图书馆进行实证评估,并提出信息安全风险控制措施^[96]。

2.3 国内数字图书馆评价存在的主要问题与对策

国内数字图书馆评价这几年取得了不少成绩,但由于存在以下几个方面的问题,使其发展受到制约。

(1) 缺乏有组织的评价研究和实践活动。目前有机组的指导性活动只有教育部高校图工委和CALIS管理中心组织的对高等学校电子资源方面的与采购相关的评价,与美国、英国、欧洲有关组织开展的大范围、广泛参与、长期积累的评价活动形成了鲜明对比。

(2) 缺乏统计标准规范。由于缺乏行业统一的评价标准,就无法制定与之相适应的统计数据标准规范,评价者也就无法获得评价所需的基本信息。例如,吴建华、王启云等开展了一些综合评价,但都遇到了评价数据难以获得的问题。行业已有的标准如COUNTER,图书馆界相关工作人员对其缺乏必要的认知,不知如何获得、解读、利用。

(3) 缺乏规范的实证评价。国内实证研究太少,除少数比较规范的研究之外,有限的评价实践普遍存在一些问题。例如采用的评价指标体系具有随意性,相互借用的情况普遍;对研究过程和具体方法、评价对象缺乏详细描述和说明;对参与评价的专家或用户的基本情况、评价时间、收集数据的数量不做任何说明;数据来源缺乏或不可靠,缺乏实践意义;对评价结果的信度和效度没有进行检验。总之,特别缺乏系统、科学、规范、用户广泛参与、长期的实证研究,小范围的评价实践科学性、严谨性也都需要加强。

(4) 相关人员缺乏评价理论和方法的训练。评价不规范、评价研究和实践重复现象严重,这也暴露出从事评价研究与实践的人员在评价理论和方面缺乏必要的训练,且需要加强对数字图书馆的认识。以可用性评价为例,新的数字图书馆或图书馆网站投入使用之前,对其进行形式化可用性评价是一个基本步骤,目的是发现可用性问题,改进设计,以更好地服务用户。但现实层面,很少有这样做的。主要原因就是设计人员缺乏对可用性的认识,也不知如何开展评价。

为此,我国需要尽快有组织地开展数字图书馆评

价研究,建立起我国自己的评价组织体系,把研究和试验评价的指标体系上升为行业或国家标准。针对形成的评价标准,建立统计数据规范和正式的数据采集渠道,最终建立评价需要的基本统计信息数据库。与此同时,加强行业培训,针对不同类型的人员如网站设计人员、资源组织和管理人员、数字图书馆规划设计人员等开展有针对性的培训,特别要在评价方法、统计分析方法等方面加强训练,让他们学会利用评价这一手段,通过科学、严谨的评价促进数字图书馆开发利用水平的提高。信息管理类专业的课程也需要加强这方面的内容。

3 结语:数字图书馆评价发展趋势

回顾近五年来国内外数字图书馆评价的研究与实践活动,可以发现其发展呈现出以下四大趋势。

第一,以用户为中心的评价和用户参与的评价。随着数字图书馆由技术中心向使用中心转移,评价的对象也从系统中心转移到用户中心。具体表现在通过用户满意的评价、反复的评价进行以用户为中心的再设计,用户参与的评价,以及建立用户视角的评价标准和开展由用户主导的评价活动。总之,评价研究与实践活动都受“以用户为中心”原则的指引。

第二,标准化和自动化评价。数字图书馆记录有大量用户和系统活动,利用日志文件建立自动评价系统被证明是可行的。通过自动采集用户行为和偏好信息,测度和评估用户和系统之间的概念错位,由此建立的自适应数字图书馆,比用户适应系统的数字图书馆能够更好地满足用户信息需求。自动评价建立在标准的基础上,只有标准化才能实现同类系统的比较评价。

第三,日志与调查相结合实现全面评价。从系统的角度看,日志信息易于获取,但有其局限性。为此,需要通过网络调查的手段获取一些通过日志无法获得的用户信息,从而实现全面评价。

第四,数字图书馆与用户高度融合挑战现有评价手段。随着数字图书馆的发展,评价对象日趋复杂。例如:数字图书馆新的检索模式,含有语义成分的数字图书馆,以数字图书馆的文献为基础进行意义建构,用户写作环境与数字图书馆的紧密集成,等等。这些新发展为用户提供了更强更丰富的功能,实现了用户与系统的高度融合,从而支持用户的日常活动。但是,这些系统的发展需要反复进行测度→评价→反馈→改进→再测度,现有的评价手段已被证明是不适应的。发展新的评价技术,是数字图书馆发展的必然要求。

参考文献

- [1] KYRILLIDOU M, GIERSCHE S. Developing the DigiQUAL Protocol for Digital Library Evaluation[C]. Denver, CO: 2005.
- [2] KYRILLIDOU M E A. DigiQUALTM: A Digital Library Evaluation Service[C]. South Africa, 2007.
- [3] FALETAR S. Libraries in the Digital Age: LIDA 2005[J]. Library Hi Tech News, 2005(8): 12-13.
- [4] SARACEVIC T. How were digital libraries evaluated?: LIBRARIES IN THE DIGITAL AGE (LIDA) 2005[Z]. Dubrovnik and Mljet, Croatia: 2005: 2010.
- [5] NICHOLSON S. The basis for bibliomining: Frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services[J]. Information Processing & Management, 2006, 42(2): 293-326.
- [6] XIE H I. Evaluation of digital libraries: Criteria and problems from users' perspectives[J]. Library & Information Science Research, 2006, 44(3): 1346-1373.
- [7] XIE H I. Users' evaluation of digital libraries (DLs): Their uses, their criteria, and their assessment[J]. Information Processing & Management, 2008, 28(3): 433-452.
- [8] ISFANDYARI-MOGHADDAM A, BAYAT B. Digital libraries in the mirror of the literature: issues and considerations [J]. The Electronic Library, 2008, 26(6): 844-862.
- [9] BERTOT J C, SNEAD J T, JAEGER P T, et al. Functionality, usability, and accessibility: Iterative user-centered evaluation strategies for digital libraries[J]. Performance Measurement and Metrics, 2006, 7(1): 17-28.
- [10] CHOWDHURY S, LANDONI M, GIBB F. Usability and impact of digital libraries: a review[J]. Online Information Review, 2006, 30(6): 656-680.
- [11] GONCALVES M A E A. "What is a good digital library?" - A quality model for digital libraries[J]. Information Processing & Management, 2007(43): 1416-1437.
- [12] MOREIRA B L E A. Evaluating Digital Libraries with 5SQual[C]. 2007.
- [13] MOREIRA B L, GONCALVES M A, LAENDER A H F, et al. Automatic evaluation of digital libraries with 5SQual[J]. Journal of Informetrics, 2009, 43(5): 589-595.
- [14] SHEN R E A. What is a successful digital library?[C]. 2006.
- [15] FUHR N. Evaluation of digital libraries[J]. International Journal of Digital Libraries, 2007, 8(1): 21-38.
- [16] MOSKOVITCH R, MD S B M, MSC, et al. A Comparative Evaluation of Full-text, Concept-based, and Context-sensitive Search[J]. Journal of the American Medical Informatics Association, 2007, 42(1): 11-21.
- [17] MOSKOVITCH R, SHAHAR Y. Vaidurya: A multiple-ontology, concept-based, context-sensitive clinical-guideline search engine[J]. Journal of Biomedical Informatics, 2009, 14(2): 164-174.
- [18] BAINBRIDGE D, DEWSNIP M, WITTEN I H. Searching digital music libraries[J]. Information Processing & Management, 2005, 20(2): 71-83.
- [19] PRICE S L, NIELSEN M L, DELCAMPRE L M L, et al. Using semantic components to search for domain-specific documents: An evaluation from the system

perspective and the user perspective[J]. Information Systems,2009,13(6):339-381.

- [20] SANTOS R L T, ROBERTO P A, GONGALVES M A, et al. A Web services-based framework for building componentized digital libraries[J]. Journal of Systems and Software,2008,33(7):699-723.
- [21] FAROOQ U, GANOE C H, CARROLL J M, et al. Design and evaluation of awareness mechanisms in CiteSeer[J]. Information Processing & Management,2008,67(4):297-312.
- [22] TWIDALE M B, GRUZZD A A, NICHOLS D M. Writing in the library: Exploring tighter integration of digital library use with the writing process[J]. Information Processing & Management,2008,44(2):556-557.
- [23] FRIAS-MARTINEZ E, CHEN S Y, LIU X. Evaluation of a personalized digital library based on cognitive styles: Adaptivity vs. adaptability[J]. International Journal of Information Management,2009,28(5):355-365.
- [24] UREN V, SHUM S B, BACHLER M, et al. Sensemaking tools for understanding research literatures: Design, implementation and user evaluation[J]. International Journal of Human-Computer Studies,2006,64(5):420-445.
- [25] GOH D H, ANQI A C D. A checklist for evaluating open source digital library software[J]. Online Information Review,2006,30(4):360-379.
- [26] JENG J H. Usability of the Digital Library: An Evaluation Model[D]. The State University of New Jersey, 2006.
- [27] ZIMMERMAN D, Professor, PASCHAL D B, et al. An exploratory usability evaluation of Colorado State University Libraries' digital collections and the Western Waters Digital Library Web sites[J]. The Journal of Academic Librarianship,2009,26(2):141-150.
- [28] NAIKO J. Anthropological motivated usability evaluation An exploration of IREON – international relations and area studies gateway[J]. Library Hi Tech,2008,28(4):606-621.
- [29] KIM H H, Kim Y H. Usability study of digital institutional repositories[J]. The Electronic Library,2008,26(6):863-881.
- [30] LONG H, LAGE K, CRONIN C. Usability testing in the context of user-centered design[J]. OCLC Systems & Services,2005,21(4):324-345.
- [31] TSAKONAS G, PAPAETHODOROU C. Exploring usefulness and usability in the evaluation of open access digital libraries[J]. Information Processing & Management,2008,44(3):1346-1373.
- [32] ZHANG X, LI Y, LIU J, et al. Effects of interaction design in digital libraries on user interactions[J]. Journal of Documentation,2008,64(3):438-463.
- [33] BUCHANAN S, SALAKO A. Evaluating the usability and usefulness of a digital library[J]. Library Review,2009,58(9):638-651.
- [34] KOMLODI A, HOU W, PREECE J, et al. Evaluating a cross-cultural children's online book community: Lessons learned for sociability, usability, and cultural exchange[J]. Interacting with Computers,2007,15(2):151-167.
- [35] HUTCHINSON H B E A. Supporting elementary-age children's searching and browsing: design and evaluation using the international children's digital library[J]. JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY,2007(11):1618-1630.
- [36] BLANDFORD A, ADAMS A, ATTFIELD S, et al. The PRET A Reporter framework: Evaluating digital libraries from the perspective of information work[J]. Information Processing & Management,2008,35(3):281-291.
- [37] BLANDFORD A, GREEN T R G, FURNISS D, et al. Evaluating system utility and conceptual fit using CASSM[J]. International Journal of Human-Computer Studies,2008,78(5):717-732.
- [38] ATILGAN D, BAYRAM O G. Perspectives on ... An Evaluation of Faculty Use of the Digital Library at Ankara University, Turkey[J]. The Journal of Academic Librarianship,2006,6(2):173-200.
- [39] ATAKAN C E A. An evaluation of the second survey on electronic databases usage at Ankara University Digital Library[J]. The Electronic Library,2008,26(2):249-259.
- [40] HARLEY D, HENKE J. Toward an Effective Understanding of Website Users: Advantages and Pitfalls of Linking Transaction Log Analyses and Online Surveys[J]. D-Lib Magazine,2007(March/April).
- [41] KRAMER E F. IUPUI image collection: a usability survey[J]. OCLC Systems & Services,2005,21(4):346-359.
- [42] ANDERSON I G. Pure dead brilliant? Evaluating The Glasgow Story digitisation project[J]. Program: electronic library and information systems,2007,41(4):365-385.
- [43] GARIBAY C, GUTIERREZ H, FIGUEROA A. Evaluation of a Digital Library by Means of Quality Function Deployment (QFD) and the Kano Model[J]. The Journal of Academic Librarianship,2010,17(3):221-234.
- [44] SHACHAF P, HOROWITZ S M, L. Virtual reference service evaluation: Adherence to RUSA behavioral guidelines and IFLA digital reference guidelines[J]. Library & Information Science Research,2008,28(4):501-520.
- [45] HUNTINGTON P, NICHOLAS D, JAMALI H R. Website usage metrics: A re-assessment of session data[J]. Information Processing & Management,2008,15(3):607-609.
- [46] MITTAL R, MAHESH G. Digital libraries and repositories in India: an evaluative study[J]. Program: electronic library and information systems,2008,42(3):286-302.
- [47] 乔欢. 数字图书馆评价研究[D]. 北京:北京大学,2005.
- [48] 徐革. 我国大学图书馆电子资源绩效评价方法及其应用研究[D]. 成都:西南交通大学,2006.
- [49] 过仕明. 数字参考咨询服务模式与质量评价研究[D]. 长春:吉林大学,2006.
- [50] 吴建华. 数字图书馆评价方法与实证研究[D]. 武汉:武汉大学,2008.
- [51] 王居平. 数字图书馆评价的理论和方法[M]. 合肥:安徽大学出版社,2008.
- [52] 吴建华. 数字图书馆评价方法[M]. 北京:科学出版社,2009.
- [53] 王启云. 高校数字图书馆建设评估研究[M]. 北京:现代教育出版社,2009.
- [54] 乔欢. 美国数字图书馆项目评价及评价项目研究[J]. 图书馆杂志,2005(2):59-63.
- [55] 乔欢. 数字图书馆评价客体检析[J]. 大学图书馆学报,2005(5):7-12,18.
- [56] 吴建华,王朝晖. 数字图书馆评价层次分析[J]. 情报科学,2009(8):1207-1213.
- [57] Jianhua W, Zhaohui W. Differentiated digital library evaluation in a hierarchical model stemming from its operational scope and complexity[J]. Chinese Journal of Library and Information Science,2009,2(4):54-70.
- [58] 吴建华,张千年. 研究型大学图书馆数字资源配置水平评价指标研究[J]. 图书馆杂志,2009(4):20-28.
- [59] 刘炜,楼向英,张春景. 数字图书馆评估研究[J]. 图书馆情报工作,2007(5):21-24.
- [60] 黄如花. 数字图书馆信息组织的评价[J]. 情报杂志,2005(7):7-9.
- [61] 南开大学中国图书馆网站评价研究组. 图书馆网站评价的基本理论问题[J]. 国家图书馆学报,2009(03):30-36.
- [62] 肖珑,李浩凌,徐成. CALIS数字资源评估指标体系及其应用指南[J]. 大学图书馆学报,2008(3):2-8,17.
- [63] 王启云. 高校数字图书馆建设评估指标体系研究[J]. 大学图书馆学报,2008(5):74-81.
- [64] 李枫林,李娜. 电子服务质量评价及其在数字图书馆中的应用[J]. 图书馆情报工作,2008(9):44-46,81.
- [65] 张炯,杨志周,周宁丽,等. 国家科学数字图书馆网络联合参考咨询服务质量控制及评价方案研究[J]. 现代图书情报技术,2005(11):30-33.
- [66] 叶肖恩. 网络环境下图书馆参考咨询服务评价探析[J]. 大学图书馆学报,2008(3):50-55.

- [67] 曾丽霞,张文德. 数字图书馆环境下著作权评估方法探讨[J]. 情报杂志,2006(9):48-50.
- [68] 张文德. 数字图书馆的知识产权评估方法研究[J]. 图书情报工作,2006(5):91-94.
- [69] 刘文梅,唐淑娟. 我国“211工程”高校数字图书馆绩效评估研究——已有研究提出的指标在我国的适用性[J]. 大学图书馆学报,2005(5):13-18.
- [70] 于良芝,等. SERQUAL与我国高校图书馆服务质量评价——关于SERQUAL适用性的定量研究[J]. 图书情报工作,2005(6):90-94.
- [71] 吕娜,余锦凤. 基于焦点团体的数字图书馆评价研究框架[J]. 图书情报工作,2009(9):29-32,73.
- [72] 吴建华,魏云. 评价活动的信息原理[J]. 图书情报工作,2008(5):54-57.
- [73] 吕娜. 数字图书馆用户访问日志模型的研究与应用[J]. 现代图书情报技术,2006(12):5-8.
- [74] 吕娜. 基于频繁下载集的数字图书馆日志挖掘实验[J]. 图书情报工作,2007(5):138-141.
- [75] 唐琼. 高校图书馆电子资源使用评价研究——以中山大学图书馆为例[D]. 广州:中山大学,2006.
- [76] 刘海霞,方平,胡德华. 开放存取期刊的质量评价研究[J]. 图书馆杂志,2006(6):23-27.
- [77] 胡德华,常小婉. 开放存取期刊论文质量和影响力的评价研究[J]. 图书情报工作,2008(2):61-64.
- [78] 陶雯,胡德华,曲艳吉,等. 开放存取期刊质量评价方法研究[J]. 图书情报工作,2006(10):72-75.
- [79] 侯集体. 开放存取期刊质量评价指标研究[J]. 图书情报工作,2009(12):140-143.
- [80] 庞弘毅,徐文贤. 数字资源评价与优选指标体系研究[J]. 图书情报工作,2009(17):27-30,125.
- [81] 毛莉,陈惠兰. 外文电子资源的评价分析——东华大学图书馆案例研究[J]. 图书馆杂志,2009(12):29-32.
- [82] 张彬. 外文电子期刊全文数据库评价模型[D]. 南京:南京农业大学,2007.
- [83] 吕娜,余锦凤. 数字图书馆以用户为中心的通用满意度模型的构建[J]. 情报学报,2006(3):322-325.
- [84] 甘利人,等. CSI模型适用性分析[J]. 数字图书馆论坛,2007(12):8-13.
- [85] 李莉,等. 基于科技数据库网站的顾客满意度指数 (ICSI) 模型构建[J]. 数字图书馆论坛,2007(12):14-19.
- [86] 李莉,等. 科技数据库网站信息用户满意度测评研究及实证分析[J]. 情报学报,2007(4):546-554.
- [87] 丁震春,李冲,李莉,等. 图书馆网站满意因素微观自主评价软件的设计[J]. 现代图书情报技术,2009(04):40-43.
- [88] 黄晓斌,卢琰. 论数字图书馆用户界面的评价[J]. 图书馆论坛,2005(6):16-19.
- [89] 向建军. 基于信息构建原则的我国高校数字图书馆可用性评价研究[D]. 重庆:西南大学,2007.
- [90] 南开大学中国图书馆网站评价研究组. 我国“211”高校图书馆网站评价[J]. 国家图书馆学报,2009(03):45-53.
- [91] 南开大学中国图书馆网站评价研究组. 我国省级公共图书馆网站评价[J]. 国家图书馆学报,2009(03):37-44.
- [92] 赵宇翔. 公共图书馆网站信息构建可用性评价研究[J]. 现代图书情报技术,2007(3):60-64.
- [93] 赵宇翔,徐一新. 高校图书馆网站信息构建的可用性研究[J]. 图书馆杂志,2009(1):48-56.
- [94] 王晶晶. 图书馆数字参考咨询服务质量评价指标体系研究[D]. 哈尔滨:黑龙江大学,2009.
- [95] 赵坚. 数字图书馆信息安全风险评估辅助工具的开发与设计[D]. 南京:南京农业大学,2008.
- [96] 朱晓欢. 基于ISO27000的复合图书馆信息安全风险评估理论与实证研究[D]. 南京:南京农业大学,2007.

作者简介

吴建华 (1964-), 华中师范大学信息管理系教授, 管理学博士。研究方向: 数字图书馆及其评价、信息测度与评价方法。发表论文20余篇, 出版专著1部, 参编著作4部。通讯地址: 湖北省武汉市珞瑜路152号华中师范大学信息管理系 430079。E-mail: wujh@mail.ccnu.edu.cn

何汶 (1965-), 武汉大学图书馆副研究馆员。研究方向: 科学评价与管理。发表论文20余篇, 参编著作4部。通讯地址: 湖北省武汉市珞珈山 武汉大学图书馆 430072

Retrospect and Prospect on Research and Practice of Digital Libraries Evaluation at Home and Abroad:2005-2009

Wu Jianhua / Department of Information Management, Huazhong Normal University, Wuhan, 430079

He Wen / Library of Wuhan University, Wuhan, 430072

Abstract: The evaluation research and practice activities of digital libraries at home and abroad are overviewed. Some typical theories, models, criteria, systems, methods, cases are introduced. The main problems of domestic evaluation are revealed and countermeasures are raised. It summarizes the main trends of digital libraries evaluation.

Keywords: Digital libraries, Evaluation, Practice, Overview

(收稿日期: 2010-02-27)