

数字图书馆交互评估：从理论构建到工具开发^{*}

□李月琳 张昕

摘要 基于数字图书馆交互的一系列前期研究及对相关研究的回顾,构建了由52个指标构成的数字图书馆交互评估问卷,开展预调查修改和完善问卷,并在正式调查中收集了372份有效问卷。研究分析了各指标的重要程度、信度及效度。采用因子分析对52个指标进行聚合,经过筛选和反复验证,最终提炼出包括系统的效用、网站结构与信息组织、资源可获得性、页面表现、可用性、主观任务感知及系统的绩效与效率在内的7个因素及45个指标,构成数字图书馆交互评估量表,为评估数字图书馆的交互提供了具备有效性、可靠性和可操作性的评估工具。同时,文章提出使用该工具的相关建议,阐述如何使用该评估工具开展数字图书馆交互评估。

关键词 数字图书馆 数字图书馆交互 评估工具

分类号 G250

DOI 10.16603/j.issn1002-1027.2018.02.009

数字图书馆评估可分为整体性评估和局部性评估。整体性评估着眼于数字图书馆的整体绩效,从数字图书馆的不同方面评估数字图书馆。如张颖(Zhang Ying)从不同的利益相关者切入,包括用户、研究者、管理者、开发者及图书馆员,探究他们对数字图书馆的内容、技术、界面、服务、用户和环境的评估标准,提出并构建了“整体性数字图书馆评估模型”^[1]。而局部性评估则着眼于数字图书馆的不同侧面,如评估数字图书馆的服务、资源、功能设计等。针对数字图书馆的交互评估是局部性评估的重要方面。“交互”是描述用户与基于计算技术的各类系统之间互动的重要概念。有效的交互应能支持用户完成与系统的一系列互动,满足用户的信息需求及对资源的具体要求,从而支持其完成搜索和工作任务。因而,支持用户与数字图书馆交互的相关设计是数字图书馆的核心技术之一。因而,本研究探讨“数字图书馆交互评估”,着重考察数字图书馆的交互功能设计对用户交互过程的支持程度。然而,虽然此前的一系列研究提出了数字图书馆交互评估的理论模型并不断细化^{[2][3]},数字图书馆环境下的交互评估如何具体实践,即可操作的评估工具的构建仍然是有待解决的问题。为此,本研究试图开发一套可用

于评估数字图书馆交互功能的工具,以完善数字图书馆的交互评估,实现从理论构建到实践应用的跨越。

1 研究背景与文献回顾

1.1 前期研究:交互绩效与交互评估

本研究建立在一系列的探讨数字图书馆环境下用户与系统交互绩效与交互功能评估的理论与实证研究的基础之上。前期研究之一探讨数字图书馆环境下影响交互绩效的因素^[4],构建了用户与数字图书馆的三维交互模型,并设计实验,探讨影响用户与数字图书馆交互绩效的因素。三维交互模型中的“三维”包括信息维、技术维和任务维。该研究进一步细化了各个维度及测量指标。同时,将用户的交互绩效界定为用户对交互成功的主观感知,如交互的成功、所获结果的有用性等。研究结果显示,用户在三个维度的交互均一定程度上显著影响用户的交互绩效。

研究之二针对信息检索研究中的任务设计^[5],探讨研究中使用的模拟仿真工作任务情境方法的适当性及如何使其发挥更好的效用。研究发现模拟仿真工作任务情境与真实工作任务情境在任务属性上

^{*} 本研究受国家社科基金项目“基于用户多维交互行为的数字图书馆可持续发展评估模型与实证研究”(11BTQ009)的支持。

通讯作者:李月琳,ORCID:0000-0002-1496-6741,yuelinli@nankai.edu.cn。

存在某些显著差异。然而,在用户反馈、交互信息搜索行为和交互绩效上均不存在显著差异。也就是说,仿真和真实工作任务的差异并不会显著影响用户与数字图书馆之间的交互,前者可以有效替代后者用于信息检索系统评估及用户信息搜寻行为的研究。同时,本研究提出合理设计模拟仿真工作任务的必要性。同时指出,针对模拟仿真工作任务的设计,除皮娅·博隆德(Pia Borlund)提出的指导性建议^[6]外,还必须清楚了解目标用户群的真实工作任务,控制任务的关键属性,如任务复杂度等。

研究之三构建了数字图书馆交互评估理论模型^{[2][3]}。该研究采用实验研究方法,招募了48名实验参加者,包括本科生、硕士研究生和博士研究生各16名,分别代表不同层次的学生群体,揭示了影响用户对数字图书馆交互功能评估的关联性指标和预测性指标,并据此构建了基于用户多维交互的数字图书馆交互功能评估的理论模型^{[2][3]}。基于这一模型,研究之四从用户视角探讨用户与数字图书馆交互的关键成功因素^[7]。研究采用实验研究方法,并在实验过程中嵌入深度访谈,以实时性地探知用户与系统交互过程中支持其交互成功或导致其交互失败的关键因素。最终确定了21个关键成功因素^[7]。

基于前期研究不断完善和细化的理论模型,本研究开发数字图书馆的交互评估工具,完成从理论构建到实践工具的开发,以期更好地服务于数字图书馆的交互评估实践及交互功能开发与设计。

1.2 相关研究综述

1.2.1 数字图书馆评估与交互研究

加里·马尔基西奥尼尼(Gary Marchionini)认为数字图书馆的评估应以用户为中心,并提出了纵向多面评估理论,包括对系统、社区、基础设施和机械优势进行评估^[8]。泰夫科·萨拉塞维克(Tefko Saracevic)将数字图书馆评估分为以系统为中心和以用户为中心两个层面,以系统为中心涵盖了工程、处理和内容层面;以用户为中心涵盖了个人、社会和机构层面^[9]。解虹(Hong I. Xie)则着重从用户视角来探讨数字图书馆评价指标,其构建的评估指标多集中于用户学习、检索信息的效率和实用性上^[10]。詹尼·察克纳斯(Giannis Tsakonas)基于对数字图书馆评估本体论(Digital Library Evaluation Ontology, DiLEO)的研究,认为当前的数字图书馆评估多以系统为中心,涵盖了绩效、效果和技术三个维

度^[11]。马西塔·阿曼达(Masitah Ahmada)和杰马尔·阿贝瓦吉(Jemal H. Abawajy)从服务角度出发构建数字图书馆评估模型^[12]。

国内学者的研究多集中于数字图书馆评估原则的构建及指标的 settings。这些原则包括:评估体系的导向性^{[13][14]}、科学性^[11]、客观性^[13]、实用性^[13]原则;评估对象的确定性原则^{[11][15]};评估标准的灵活性和通用性原则^{[1][11][13]};评估手段的技术性原则^{[1][11][13]}。评估指标的 settings 大多从宏观角度出发,包括数字图书馆的馆藏资源^[16]、用户及使用^[15]、网站建设^{[1][13][15]}、价值与成本计算^[13]、服务^{[1][13][17]}等,这些宏观维度的确定为评估研究奠定了一定的基础。在研究方法上,多利用层次分析法^{[18][19][20][21]}、德尔菲法^[13]和模糊综合评价法^[22]等。这些研究方法一般用于数字图书馆的整体性评估,通过计算分值来确定指标权重或者比较不同数字图书馆的质量水平。

此外,国内学者日益重视从用户角度进行数字图书馆评估^[23]。夏立新等从用户满意度视角出发,建立了基于数字图书馆设施、流程与内容质量的评估模型^[18]。徐芳、戴炜轶和赵杨从用户体验角度出发对数字图书馆质量评估进行探讨。前者通过设计评估指标,对国内数字图书馆进行比较,但缺少评估体系的理论化构建^[24];后者构建了用户体验控制框架,但模型构建未经过实证研究的检验^[25]。另有一些学者基于用户视角从界面^[27]、信息构建^[28]等角度探讨评估问题。

在数字图书馆服务评估中,用户交互是非常重要的^[29]。除以上研究探讨了数字图书馆环境下的用户与系统的交互绩效及交互评估研究之外,诺伯特·福尔(Norbert Fuhr)等构建了包括用户、内容和系统之间交互关系的三维模型,揭示了:①用户—内容关系与有用性因素相关;②用户—系统关系与可用性因素相关,包括满意度、效果等;③内容—系统关系与绩效因素相关,包括响应时间、查全率、查准率等^[30]。一些学者从用户交互角度出发^{[31][32]},构建了基于交互需求、交互过程、交互环境的评估模型。

1.2.2 数字图书馆评估工具

马科斯·贡萨尔维斯(Marcos A. Goncalves)等从5S(Streams, Structures, Spaces, Scenarios and Societies)模型中抽取了数字图书馆的高阶(High—

level)概念,总结出与之相对应的质量维度,对于每一个质量维度,都由基于 5S 和已知数字图书馆概念的集合和函数来表示,形成了一套可供量化的评估工具^[33]。芭芭拉·莫雷拉(Bárbara L. Moreira)等以贡萨尔维斯设计的评估工具为基础,设计旨在帮助数字图书馆管理者在数字图书馆建设和维护阶段可重复使用的简捷的质量评估工具—5SQual^[34]。欧盟得洛斯(Delos)工作小组根据其提出的数字图书馆概念模型,建立了整体评估工具。此外,研究图书馆学会(Associate of Research Libraries)建立了电子资源计量(E-Metrics)指标评估工具,旨在测度数字图书馆对于电子信息资源和服务的投资是否取得了收益、是否创造了更高的用户使用率^[35]。

数字图书馆服务评估工具的开发受到众多学者的关注。皮特·赫农(Peter Hernon)等学者最早将电子服务质量评估量表引入到图书馆领域,他们借鉴 SERVQUAL 和 E-S-QUAL 等评估量表的概念理论和研究方法,对数字图书馆服务的质量以及影响因素进行深入分析,建立了图书馆服务质量评价模型(Library E-servqual)^[36]。基兰·考尔(Kiran Kaur)等学者在此研究的基础上,构建了一个三层体系的数字图书馆服务质量评估量表,从环境质量、传递质量和结果质量三个方面进行了评估^[37]。国内学者,如施国洪等在 SERVQUAL 量表的基础上,建立了图书馆服务质量测评量表^[38]。但是其测量针对传统图书馆。郭瑞芳^[39]、文思敏^[40]分别参考 LibQUAL+TM 量表构建了移动图书馆服务质量评估体系。

可见,数字图书馆领域更多关注数字图书馆的整体或服务质量的评估工具或量表的开发,针对数字图书馆交互功能的评估工具相对较少;另一方面,对于数字图书馆的交互评估而言,目前多为模型框架或者规范指南,缺乏可操作的量表,难以在实践中推广应用。因此,本研究试图弥补这一缺憾,基于前期的一系列理论研究,开发数字图书馆交互评估工具,以支持数字图书馆交互的评估实践。

2 研究方法

本研究采用问卷调查,以数字图书馆交互功能评估模型为理论基础、交互的关键成功因素为框架,设计数字图书馆交互评估问卷,并通过分析指标重要性,开展因子分析及信度、效度分析。

2.1 问卷设计

问卷设计基于前期研究得出的数字图书馆交互三维模型、交互评估理论模型及关键成功因素模型,分为人口统计变量部分和主体部分。主体部分包括 53 个题项,考察从用户视角测量不同指标对用户评估数字图书馆交互功能支持程度的重要性。其中题项 1—52 采用李克特七点量表,“1”表示非常不重要,“7”表示非常重要。同时,用“NA”表示该指标对用户来说不适合评价数字图书馆的交互功能,“NA”列在“7”之后。

预调查共发放问卷 28 份,同时要求被试者对问卷进行评价,提出建议。预调查中受访男性占比为 46.4%,女性占比为 53.6%,男女比例大体均衡。从学历来看,涵盖了不同层级的学生,基本反映了不同学历层次受访者的情况。根据数据分析结果及受访者的反馈,对问卷中的语言表述进行修改,最终形成本研究的正式问卷(本文略)。

2.2 问卷发放

由于数字图书馆的主要用户集中在高校,本研究的调查主要以高校本科生、硕士研究生以及博士研究生为对象,集中在当地两所教育部直属高校进行发放。问卷要求受调查者当面填写完成,发放地点为学生活动的主要场所,如自习室等。

本研究共发放问卷 400 份,经过筛选,剔除无效问卷,获得 372 份有效问卷,问卷有效率为 93%。本研究收集的数据采用 SPSS22.0 进行处理。

2.3 样本的人口统计学变量特征

样本的人口统计学基本特征如表 1 所示:

表 1 样本特征

描述性统计指标	选项	样本数量 (百分比)	所占百分比
性别	男	174	46.8%
	女	198	53.2%
年龄	<18 岁	3	0.8%
	18—24 岁	266	71.5%
	25≤—29 岁	81	21.8%
	≥30 岁	22	5.9%
学历	本科生	153	41.1%
	硕士研究生	127	34.1%
	博士研究生	92	24.7%

描述性统计指标	选项	样本数量 (百分比)	所占百分比
学科①	人文艺术学科	35	9.3%
	社会科学	158	42%
	理工医学科	183	48.7%

从基本信息来看,受调查者的性别比例均衡,年龄集中在18—25岁,共266人,占总样本的71.5%,反映了高校学生的总体特征。从学历来看,各层次学生占比合理;从学科来看,以社会科学和理工科的学生占比居多,其他学科也有少量分布。样本具有一定的代表性。

2.3 样本的数字图书馆经验

为进一步探知样本的有效性,本研究从用户使用数字图书馆的频率、最常使用的数字图书馆、使用的方式、目的,信息检索的专业程度及检索水平、效果等方面揭示受调查者的特征,结果如图3—7及表6所示:

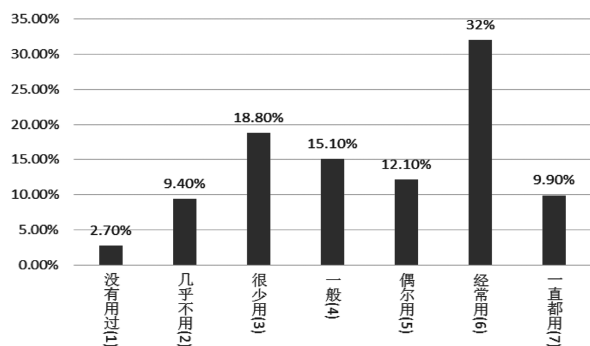


图3 使用数字图书馆的频率

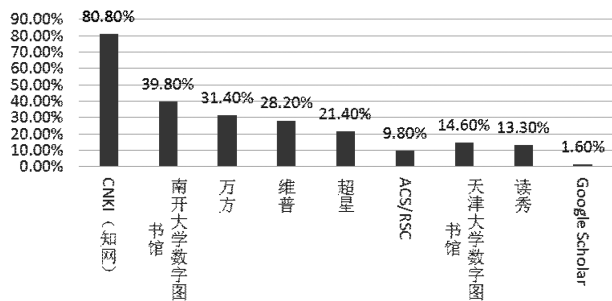


图4 最常使用的数字图书馆

(自我报告题项,提供的数字图书馆可不唯一)

图3显示,经常使用数字图书馆的受调查者(包括经常使用和一直使用)为156人,所占百分比为41.9%,占比最高。图4展示了受调查者最常使用的数字图书馆,其中CNKI(知网)、南开大学数字图书馆、万方、维普、超星位居前五。

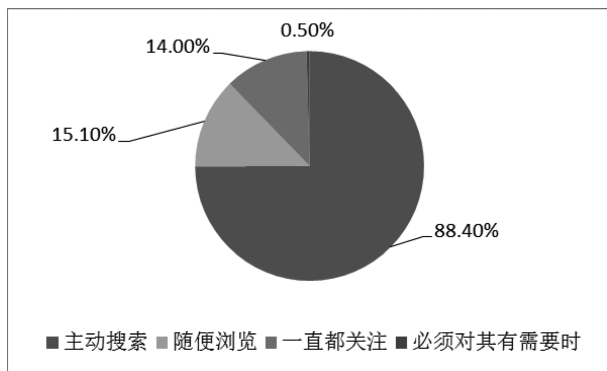


图5 使用数字图书馆的方式(可多选)

图5显示,大多数受调查者使用数字图书馆的方式为主动搜索,占比高达88.4%;表2表明,使用数字图书馆的目的多为撰写论文、科研需要以及完成课堂作业;另外,64.5%的受调查者并未学习过相关课程。

表2 使用数字图书馆的目的(可多选)

选项	样本数量	百分比
解决工作中问题	109	29.3%
撰写论文	286	76.9%
完成课堂作业	186	50.0%
科研需要	226	60.8%
学习需要	1	0.3%
解决生活中问题	35	9.8%
没什么目的	17	4.6%

图6显示,52.4%的受调查者认为自己的检索水平一般,占比最高;图7表明,46.5%的受调查者认为自己的检索效果一般,占比最高。其次,20.2%的受调查者认为自己的检索水平比较高(见图6),42.5%认为自己的检索效果比较成功(见图7)。可见,多数的受调查者认为自己的检索水平一般或较好,检索比较成功。

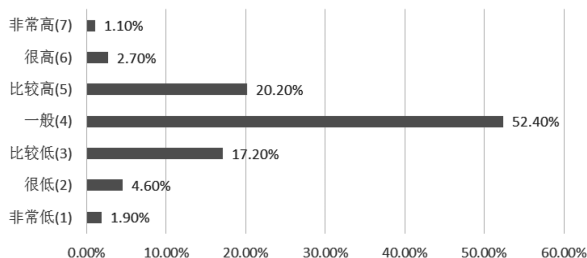


图6 样本的检索水平

① 因为部分受访者为双修,所以选择了两个学科,统计的时候样本量多于有效问卷数。

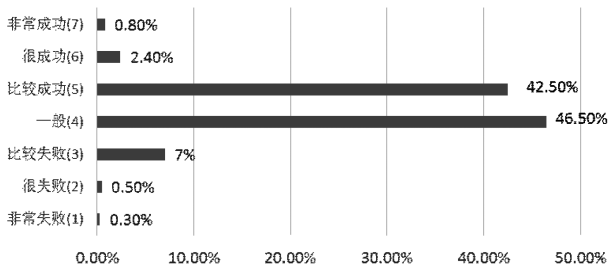


图7 样本的检索效果

数据分析表明,受调查者一定程度上了解数字图书馆,学习、工作过程中使用数字图书馆,并具有检索经验。

3 研究结果

3.1 指标的重要程度分析

本研究选取的 52 项指标均来源于前期研究逐步构建并细化的理论模型,为进一步验证这些理论模型的可靠性及用于评估数字图书馆交互的可能性,本研究要求受调查者评价这些指标在评估数字图书馆交互中的重要程度。结果表明,各指标的均值分布在 3.76 与 5.88 之间。除 Q15 的标准差为 3.93 外,其他的各指标的标准差分布在 1.05 和 1.46 之间。部分指标的重要程度见表 4 和表 6。由此可见,受调查者对这些指标在评估数字图书馆交互功能方面的认同度较高。除 Q15 之外,其他指标均能取得一定程度的共识。因而,以下针对所有指标进行进一步分析。

3.2 信度、效度检验

由于本问卷中的所有指标均用于测量数字图书馆交互功能,因而,信度检验针对问卷中的所有题项。检验结果显示克隆巴赫系数为 0.95,说明问卷信度非常高。

利用因子分析中的主成份分析法测量问卷中的原始数据,可以对数据的效度进行有效检验,达到降维的作用,也可以发现各个指标之间的关系。分析表明,KMO 检验值为 0.943。Bartlett 球形检验近似卡方值为 10455.586, $p < 0.01$,样本适合因子分析。

采用因子分析得出问卷指标相关矩阵的特征值以及累计贡献率,一共提取出 10 个公因子,累计贡献率为 62.932%,表明问卷的效度较好。

3.3 公因子提取

利用方差最大化的正交旋转法,定义特征值大

于 1、因子载荷大于 0.5 以上的指标,对问卷中的 52 个指标进行因子分析,旋转后得到了因子载荷矩阵,如表 3 所示。

表3 第一次因子分析结果(表中 DL 为数字图书馆的英文缩写)

公因子	评估指标	因子载荷
F1	Q49 我在 DL 搜寻信息的顺利程度	0.766
	Q47 我能够在 DL 界面上进行高效率的信息搜索	0.761
	Q48 我能够在 DL 界面上进行高效率的信息浏览	0.707
	Q46 我能够方便地操控 DL 实现自己的搜寻目标	0.671
	Q51 通过 DL 搜寻到的结果令我满意的程度	0.633
	Q45 DL 的各种功能是容易学习和使用的	0.623
	Q50 在 DL 搜寻信息的过程中,系统的表现令我满意的程度	0.618
	Q43DL 提供的信息内容真实无误	0.603
	Q44 我在 DL 里搜索不同内容的信息时均能得到满足	0.567
	Q42 我从 DL 中获得的信息都是有用的信息	0.551
F2	Q52 在 DL 搜索的过程中我要付出的努力程度	0.492
	Q22DL 的资源分类是合理的	0.708
	Q20DL 对相关网站有详细的分类链接	0.698
	Q21 DL 的相关网站都是有效链接	0.690
	Q23DL 链接的层级数量合适	0.642
	Q40 DL 对各类资源的区分有明确的标识	0.551
	Q24DL 的链接能够引导我方便查找信息	0.550
	Q26 顶部或底端的导航条十分容易找到	0.514
	Q41DL 提供的搜索关键词准确	0.487
	Q25DL 网站的反应速度快,稳定性好	0.437
F3	Q10DL 提供了完整的目录	0.412
	Q36 DL 的功能设置集中,方便使用	0.335
	Q28 页面排版和布局合理	0.720
	Q29DL 符合网站通用的设计规则	0.673
	Q30 网页整洁有序、框架脉络清晰,对搜索很有帮助	0.656
	Q31 搜索过程中,界面层次安排符合我平常的浏览习惯	0.652
	Q27 搜索工具,如检索框、作者、主题等索引容易找到	0.541
	Q32 我能从 DL 链接到其他检索系统/数据库	0.437

公因子	评估指标	因子载荷
F4	Q1 DL 里资源的数量能满足我信息获取的需要	0.836
	Q2 DL 里资源的种类能满足我信息获取的需要	0.799
	Q3 DL 里资源的内容质量能满足我信息获取的需要	0.780
	Q4 DL 提供了方便的下载方式	0.515
	Q5 DL 检索结果与我想要的结果相关性高	0.432
F5	Q17 DL 提供了能够满足需要的多种功能	0.689
	Q18 页面上主题的字体和大小能够吸引我的注意力	0.635
	Q16 DL 网站呈现的内容是详细的	0.621
	Q12 DL 提供了完备的用户指南	0.598
	Q11 DL 的资源分类是详尽的	0.528
F6	Q38 当有检索任务时,我有信心利用 DL 完成任务	0.796
	Q37 当有检索任务时,我对 DL 足够熟悉以至于能帮助我很好地完成任务	0.770
	Q39 当有检索任务时,我所拥有的专业知识能够帮助我更好地使用 DL	0.756
	Q34 利用 DL 之前,我通常已经明确了搜索目标	0.437

公因子	评估指标	因子载荷
F7	Q6 在 DL 检索时我可以通过多种途径链接到同一页面	0.669
	Q7 DL 里资源的更新速度快	0.516
	Q9 DL 页面上的主题能准确表达文章或词条的核心内容	0.449
	Q8 DL 能够提供与检索需求相关的内容	0.436
	Q14 DL 对用户的查询要求反应迅速	0.372
F8	Q33 DL 有适度的网站宣传	0.512
	Q35 相比于其他数据库,我更依赖于我所选择的 DL	0.503
	Q13 即使初次使用 DL,也易于学习,使用起来并不困难	0.437
F9	Q15 DL 页面的内容简洁	0.797
F10	Q19 页面上主题的颜色能够吸引我的注意力	0.449

表 3 显示,指标 Q5、Q8、Q9、Q10、Q13、Q14、Q19、Q25、Q32、Q34、Q36、Q41 和 Q52 的因子载荷低于 0.5,把上述 13 个指标删除后,将剩余因子载荷在 0.5 以上的指标进行第二次因子分析,累计贡献率为 61.608%,一共得到 7 个公因子,见表 4。由于第二次因子分析析出的因子极有可能成为构建量表的因子和指标,表 4 同时列出了这些指标的重要程度均值及标准差。

表 4 第二次因子分析析出的公因子及重要程度(表中 DL 为数字图书馆的英文缩写)

公因子	指标	因子载荷	均值(标准差)
F1	Q49 我在 DL 搜寻信息的顺利程度	0.759	5.63(1.05)
	Q47 我能够在 DL 界面上进行高效率的信息搜索	0.739	5.73(1.06)
	Q43DL 提供的信息内容真实无误	0.708	5.88(1.27)
	Q48 我能够在 DL 界面上进行高效率的信息浏览	0.704	5.70(1.11)
	Q51 通过 DL 搜寻到的结果令我满意的程度	0.664	5.69(1.16)
	Q44 我在 DL 里搜索不同内容的信息时均能得到满足	0.653	5.41(1.25)
	Q42 我从 DL 中获得的信息都是有用的信息	0.650	5.50(1.25)
	Q50 在 DL 搜寻信息的过程中,系统的表现令我满意的程度	0.642	5.42(1.21)
	Q46 我能够方便的操控 DL 实现自己的搜寻目标	0.628	5.54(1.15)
	Q45DL 的各种功能是容易学习和使用的	0.619	5.45(1.13)
F2	Q21 DL 的相关网站都是有效链接	0.729	5.54(1.22)
	Q22 DL 的资源分类是合理的	0.717	5.55(1.08)
	Q20DL 对相关网站有详细的分类链接	0.715	4.99(1.27)
	Q23DL 链接的层级数量合适	0.652	5.23(1.19)
	Q24DL 的链接能够引导我方便查找信息	0.544	5.69(1.15)
	Q40DL 对各类资源的区分有明确的标识	0.488	5.27(1.20)

公因子	指标	因子载荷	均值(标准差)
F2	Q26 顶部或底端的导航条十分容易找到	0.476	5.23(1.28)
	Q11 DL 的资源分类是详尽的	0.459	5.44(1.22)
F3	Q1 DL 里资源的数量能满足我信息获取的需要	0.795	5.65(1.28)
	Q2 DL 里资源的种类能满足我信息获取的需要	0.787	5.60(1.32)
	Q3 DL 里资源的内容质量能满足我信息获取的需要	0.741	5.83(1.23)
	Q4 DL 提供了方便的下载方式	0.646	5.70(1.10)
	Q7 DL 里资源的更新速度快	0.471	5.62(1.34)
F4	Q28 页面排版和布局合理	0.752	5.17(1.17)
	Q29DL 符合网站通用的设计规则	0.695	5.02(1.25)
	Q31 搜索过程中,界面层次安排符合我平常的浏览习惯	0.668	5.09(1.25)
	Q30 网页整洁有序、框架脉络清晰,对搜索很有帮助	0.658	5.40(1.09)
	Q27 搜索工具,如检索框、作者、主题等索引容易找到	0.515	5.72(1.08)
F5	Q18 页面上主题的字体和大小能够吸引我的注意力	0.642	4.40(1.46)
	Q17 DL 提供了能够满足需要的多种功能	0.628	5.28(1.28)
	Q33 DL 有适度的网站宣传	0.615	4.20(1.46)
	Q16 DL 网站呈现的内容是详细的	0.602	5.31(1.29)
	Q12 DL 提供了完备的用户指南	0.580	4.88(1.46)
	Q6 在 DL 检索时我可以通过多种途径链接到同一页面	0.432	4.79(1.36)
F6	Q38 当有检索任务时,我有信心利用 DL 完成任务	0.804	5.10(1.36)
	Q39 当有检索任务时,我所拥有的专业知识能够帮助我更好地使用 DL	0.773	5.33(1.33)
	Q37 当有检索任务时,我对 DL 足够熟悉以至于能帮助我很好地完成任务	0.766	5.37(1.33)
	Q35 相比于其他数据库,我更依赖我选择的 DL	0.468	5.09(1.61)
F7	Q15 DL 页面的内容简洁	0.816	5.36(3.93)

表 4 显示,F7 仅有一个指标,其重要程度的均值虽为 5.36,但标准差相比之下偏高,为 3.93,表明数据的波动较大,受调查者对该指标的认同度存在一定的不一致性。因而,该指标可考虑剔除。

将之前所有剔除的因子再次进行主成份分析,结果如表 5 所示:

表 5 被剔除因子的第一次主成份分析结果

公因子	指标	因子载荷
M8	Q5 DL 检索结果与我想要的结果相关性高	0.815
	Q9 DL 页面上的主题能准确表达文章或词条的核心内容	0.777
	Q8 DL 能够提供与检索需求相关的内容	0.714
	Q41 DL 提供的搜索关键词准确	0.655
	Q25 DL 网站的反应速度快,稳定性好	0.586
	Q14 DL 对用户的查询要求反应迅速	0.579
	Q10 DL 提供了完整的目录	0.539
	Q13 即使初次使用 DL,也易于学习,使用起来并不困难	0.425

公因子	指标	因子载荷
M9	Q32 我能够从 DL 链接到其他检索系统/数据库	0.715
	Q36 DL 的功能设置集中,方便使用	0.642
	Q34 利用 DL 之前,我通常已经明确了搜索目标	0.609
	Q19 页面上主题的颜色能够吸引我的注意力	0.469
	Q52 在 DL 搜索的过程中我要付出的努力程度	0.384

表 5 显示,被剔除的这些指标重新聚合成两个因子,但是指标 Q13、Q19、Q52 的因子载荷低于 0.5,予以剔除。再次进行主成份分析,结果如表 6 所示。考虑到这些指标作为量表指标的可能性,表 6 同时报告了这些指标的重要程度均值及标准差。

表 6 显示,将指标 Q13、19、52 剔除后,得到的结果比较满意,所有因子的载荷都在 0.5 以上,并且聚合成两个因子,命名为 M8 和 M9。

表 6 被剔除因子的第二次主成份分析结果及重要程度

公因子	指标	因子载荷	均值(标准差)
M8	Q5 DL 检索结果与我想要的结果相关性高	0.799	5.88(1.20)
	Q9 DL 页面上的主题能准确表达文章或词条的核心内容	0.788	5.85(1.07)
	Q8 DL 能够提供与检索需求相关的内容	0.714	5.81(1.11)
	Q41 DL 提供的搜索关键词准确	0.664	5.71(1.21)
	Q25 DL 网站的反应速度快,稳定性好	0.590	5.75(1.10)
	Q10 DL 提供了完整的目录	0.590	5.27(1.30)
	Q14 DL 对用户的查询要求反应迅速	0.588	5.38(1.36)
M9	Q32 我能够从 DL 链接到其他检索系统/数据库	0.781	5.27(1.26)
	Q36 DL 的功能设置集中,方便使用	0.660	5.33(1.21)
	Q34 利用 DL 之前,我通常已经明确了搜索目标	0.652	5.52(1.38)

3.4 公因子的信度分析

将第二次因子分析得到的 6 个公因子和第二次主成份分析得到的 2 个公因子进行信度分析,结果如表 7 所示:

表 7 公因子信度分析

公因子	克隆巴赫系数
F1	0.913
F2	0.886
F3	0.855
F4	0.825
F5	0.764
F6	0.776
M8	0.814
M9	0.574

表 7 显示,仅有 M9 的克隆巴赫系数为 0.574,小于 0.7,所以删除该公因子,最终获得评估数字图书馆交互功能的 7 个公因子。

3.5 公因子命名及评估量表的形成

因子命名的过程体现了对聚合在同一因子下的指标的概念抽象,是通过提炼具体指标形成上位概念的过程。以下公因子的命名经头脑风暴和多次讨论形成,具体命名如表 8 所示。

基于以上研究结果,将问卷中各指标的表述规范化,即将所有的表述改为正向的评估,采用 7 点李克特量表,以“1”表示“完全不同意”(ED),“4”表示“中立”(N),“7”表示“完全同意”(EA),进一步开

发评估量表,见表 9。

表 8 公因子命名表

公因子	名称
F1	系统的效用
F2	网站结构与信息组织
F3	资源可获得性
F4	页面表现
F5	易用性
F6	主观任务感知
M8	系统的绩效与效率

4 讨论与结论

本研究采用问卷调查,共收集了 372 份有效问卷,通过分析指标的重要性、因子分析及信度分析,形成一套包含 7 个维度、45 项指标的数字图书馆交互评估工具。

4.1 数字图书馆交互评估维度与指标

本研究的结果显示,数字图书馆交互评估可从 7 个维度展开,即系统的效用、网站结构与信息组织、资源可获得性、页面表现、可用性、主观任务感知及系统的绩效与效率。这些维度的获得是通过因子分析聚合后形成的结果,客观地反映了用户在评估数字图书馆交互过程中所倚重的不同方面,可帮助我们更有效和准确地评估数字图书馆的交互。各项指标的设定基于前期研究的发现,通过本研究的问卷调查,再一次验证这些指标在评价数字图书馆交互中的有效性和可靠性。研究的结果支持了前期的研究发现,反映了数字图书馆的交互与技术维、信息

表 9 数字图书馆交互功能评估量表

评估维度 (D)	评估指标 (I)	ED	N	EA				
D1:系统的效用	I1:我在 DL 搜寻信息的过程顺利	1	2	3	4	5	6	7
	I2:我能够在 DL 界面上进行高效率的信息搜索	1	2	3	4	5	6	7
	I3: DL 提供的信息内容真实无误	1	2	3	4	5	6	7
	I4:我能够在 DL 界面上进行高效率的信息浏览	1	2	3	4	5	6	7
	I5:通过 DL 搜寻到的结果令我满意	1	2	3	4	5	6	7
	I6:我在 DL 里搜索不同内容的信息时均能得到满足	1	2	3	4	5	6	7
	I7:我从 DL 中获得的信息都是有用信息	1	2	3	4	5	6	7
	I8:在 DL 搜寻信息的过程中,系统的表现令我满意	1	2	3	4	5	6	7
	I9:我能够方便的操控 DL 实现自己的搜寻目标	1	2	3	4	5	6	7
	I10:DL 的各种功能是容易学习和使用的	1	2	3	4	5	6	7
D2:网站结构与信息组织	I11:DL 的相关网站都是有效链接	1	2	3	4	5	6	7
	I12:DL 的资源分类是合理的	1	2	3	4	5	6	7
	I13:DL 对相关网站有详细的分类链接	1	2	3	4	5	6	7
	I14:DL 链接的层级数量合适	1	2	3	4	5	6	7
	I15:DL 的链接能够引导我方便查找信息	1	2	3	4	5	6	7
	I16:DL 对各类资源的区分有明确的标识	1	2	3	4	5	6	7
	I17:顶部或底端的导航条十分容易找到	1	2	3	4	5	6	7
	I18:DL 的资源分类是详尽的	1	2	3	4	5	6	7
D3:资源可获得性	I19:DL 的资源的数量能满足我信息获取的需要	1	2	3	4	5	6	7
	I20:DL 的资源种类能满足我信息获取的需要	1	2	3	4	5	6	7
	I21:DL 的资源的内容质量能满足我信息获取的需要	1	2	3	4	5	6	7
	I22:DL 提供了方便的下载方式	1	2	3	4	5	6	7
	I23:DL 的资源的更新速度快	1	2	3	4	5	6	7
D4:页面表现	I24:页面排版和布局合理	1	2	3	4	5	6	7
	I25:DL 符合网站通用的设计规则	1	2	3	4	5	6	7
	I26:搜索过程中,界面层次安排符合我平常的浏览习惯	1	2	3	4	5	6	7
	I27:网页整洁有序、框架脉络清晰,对搜索很有帮助	1	2	3	4	5	6	7
	I28:搜索工具,如检索框、作者、主题等索引容易找到	1	2	3	4	5	6	7
D5:可用性	I29:页面上主题的字体和大小能够吸引我的注意力	1	2	3	4	5	6	7
	I30:DL 提供了能够满足需要的多种功能	1	2	3	4	5	6	7
	I31:DL 有适度的网站宣传	1	2	3	4	5	6	7
	I32:DL 网站呈现的内容是详细的	1	2	3	4	5	6	7
	I33:DL 提供了完备的用户指南	1	2	3	4	5	6	7
	I34:在 DL 检索时我可以通过多种途径链接到同一页面	1	2	3	4	5	6	7
D6:主观任务感知	I35:当有检索任务时,我有信心利用 DL 完成任务	1	2	3	4	5	6	7
	I36:当有检索任务时,我所拥有的专业知识能够帮助我更好地使用 DL	1	2	3	4	5	6	7
	I37:当有检索任务时,我对 DL 足够熟悉以至于能帮助我很好地完成任务	1	2	3	4	5	6	7
	I38:相比于其他数据库,我更依赖我选择的 DL	1	2	3	4	5	6	7

评估维度 (D)	评估指标 (I)	ED			N			EA		
D7:系统的绩效与效率	I39:DL 检索结果与我想要的结果相关性高	1	2	3	4	5	6	7		
	I40:DL 页面上的主题能准确表达文章或词条的核心内容	1	2	3	4	5	6	7		
	I41:DL 能够提供与检索需求相关的内容	1	2	3	4	5	6	7		
	I42:DL 提供的搜索关键词准确	1	2	3	4	5	6	7		
	I43:DL 提供了完整的目录	1	2	3	4	5	6	7		
	I44:DL 网站的反应速度快,稳定性好	1	2	3	4	5	6	7		
	I45:DL 对用户的查询要求反应迅速	1	2	3	4	5	6	7		

维和任务维三个维度密切相关。这一方面表明前期研究提出的理论模型或框架较为科学,可以对实践层面的应用进行指导;另一方面说明本研究的问卷设计和数据可靠性较高,能够较全面地反映用户对数字图书馆交互的认知与评估。另外,本研究构建的评估工具一定程度上反映了解虹(Hong I.Xie)^[10]和张颖(Zhang Ying)^[1]提出的理论模型所涵盖的部分指标,尽管因为研究的目的不同,侧重点并不完全一致。

研究结果同时启示我们,交互是一个复杂的概念,尤其从用户的视角,不仅涉及用户与系统的信息交换与互动,同时兼顾用户自身对任务的感知及交互过程的体验。研究揭示的7个维度也反映了交互这一重要概念的不同维度,不仅可帮助我们更深入地理解交互的概念,还同时启示可从不同的切入点理解“交互”。

此外,本研究通过因子分析删除了因子载荷较低的6项指标,分别是:Q13(即使初次使用DL,也易于学习,使用起来并不困难)、Q19(页面上主题的颜色能够吸引我的注意力)、Q32(我能从DL链接到其他检索系统/数据库)、Q34(利用DL之前,我通常已经明确了搜索目标)、Q36(DL的功能设置集中,方便使用)以及Q52(在DL搜索的过程中我要付出的努力程度)。这些指标的重要性程度并不低(括号中依顺序分别为均值和标准差):Q13(5.32, 1.21), Q19(3.76, 1.39), Q32(5.27, 1.26), Q34(5.52, 1.38), Q36(5.33, 1.21), Q52(5.30, 1.27)。排除这些指标的主要原因是因为本研究设定的因子载荷偏高(≥ 0.5),使它们无法聚类到任何维度中;另一方面,也不排除问题表述不够清楚合理或者是有数据异常值存在的可能。此外,排除这些指标也是为了使评估工具更加聚焦、有效。然而,这些方面也同样是不可忽视的,虽未能进入评估量表,但作为参考指

标,依然可以发挥作用。

本研究基于前期研究成果,设计52项评估指标,最终在评估量表中保留了45项,剔除指标偏少也一定程度上支持了前期研究的理论模型或框架的可靠性。

4.2 数字图书馆交互功能评估量表的使用

经过一系列研究构建的数字图书馆交互功能评估量表,适用于评估数字图书馆交互,以探知其交互功能是否支持用户与数字图书馆的交互过程,获得良好的交互绩效。该量表可用于比较不同的数字图书馆在交互功能设计方面的优劣,但这并非其主要目的。更为重要的目的在于明确待评估的数字图书馆在交互功能设计上哪些方面需要进一步改善,以帮助开发设计人员进一步完善其交互功能设计,尤其是检索和浏览功能的设计。该量表可以嵌入数字图书馆中,打造成网络问卷形式,请用户自行填写完成。如能长期收集数据,对数字图书馆的交互功能的了解和改进无疑是极有助益的。采用这样的方法,可以使数字图书馆的研究由常规性的截面研究过渡到历时性研究,以支持数字图书馆的可持续发展。也可组织专项的数字图书馆交互评估,利用该量表,通过抽样调查,完成数字图书馆评估的数据收集和分析。

如何获知妨碍数字图书馆交互过程和影响交互绩效的不同方面?可采取以下两种方式:

(1)获得适当的数据后(建议中型样本,200以上),针对每一个维度,分别考察7个维度的均值,如某维度均值低于5,该维度应引起关注并进一步考察每一项评估指标的均值,低于5的评价指标应考察其原因并针对性地加以改善和提升。也可直接考察每一项指标的均值,针对均值小于5的指标作进一步分析,以改善其设计。例如,针对页面表现维度的指标“搜索过程中,界面层次安排符合我平常的浏

览习惯”,如均值小于5,数字图书馆的开发和设计者便应重新考量页面内容的层次安排,进一步考察目标用户的浏览习惯,使其设计更符合用户的行为习惯,以提升其交互绩效。

(2)此外,获得适当数据后,针对每一维度,还可考察各分值的百分比。可事先设定基准比例进行评估,具体比例以开发方或使用方对数字图书馆的期望目标而定,可定为40%,也可为30%或20%。一旦某维度的分值低于5的比例高于基准比例,该维度便值得关注。例如,设定基准比例为30%,针对维度D1,各指标低于5的比例的均值为40%,那么该维度应该给予关注。进一步考察单一指标,着重探究分值低于5的比例高于基准比例的指标,以探知导致该比例的原因。当然,直接考察各指标的分值低于5的比例也是可行的方法,如样本中选择低于5的比例高于基准比例,其原因值得进一步考察,以寻找合适的方法和途径改善数字图书馆的交互功能设计。

当然,基于该量表,并非所有的不够理想的指标都能直接找到解决方案,如指标“我在DL搜寻信息的过程顺利”的分值低于5的比例如超过基准比例,是无法直接获得交互功能的完善方案的。需进一步研究,如采用跟踪访谈的方法,访谈部分受调查者,以更准确获知导致搜索过程不顺利的根本原因,以便对症下药地完善数字图书馆的交互功能设计。

总之,该量表提供了一个可行的、比较灵活的评估数字图书馆交互功能的方法和途径,为数字图书馆交互功能的持续改善提供实证基础。

4.3 研究的意义及启示

本研究的意义在于:①将数字图书馆交互功能评估从理论研究推向实践应用层面,结合问卷调查法和因子分析法,通过数据的具体分析构建评估工具。相比于已有的一些相关研究仍停留在理论模型构建阶段,本研究更有助于推进评估的实践。②本研究所构建的评估工具可对数字图书馆多个维度进行测量,涵盖技术维度、信息维度、任务维度及用户层面,实现对数字图书馆交互的多维评估。可见,本研究延伸了数字图书馆的评估研究,所构建的评估工具着重于数字图书馆交互层面,丰富了数字图书馆评估的内涵和方法。

评估工具的维度“系统的效用”体现了用户对数字图书馆交互功能的理解,用户满意的数字图书馆

交互功能应能有效地支持和改善用户搜寻信息的顺利程度,提高搜索和浏览的效率,提供真实及有用的信息,提供满意的搜索结果及满足用户的需求等。这些方面作为用户评估数字图书馆交互功能的重要方面,也同时启示我们数字图书馆的交互功能设计应能有效地支持用户交互才能提升用户的满意程度。“网站结构与信息组织”维度则体现了信息构建的重要性,链接的层级、链接的有效性、资源的分类及组织均能影响用户对数字图书馆交互功能的评估。同时,该维度与“页面表现”及“可用性”启示我们数字图书馆界面设计该如何更好地满足用户与数字图书馆的交互需求,如界面的层次安排、网页的整洁有序、页面排版布局等。“资源可获得性”体现了信息内容丰富对支持用户与数字图书馆交互的重要性;“系统的绩效与效率”也要求数字图书馆在资源建设方面加以重视,尤其要了解目标用户的需求及工作任务,用户对任务的主观感受也同样是用户评估数字图书馆交互支持的重要方面。可见,本研究不仅提供了可用于评估数字图书馆交互的工具,同时,也为数字图书馆的交互设计提供了思路,可帮助数字图书馆的设计者更好地完善数字图书馆的交互设计,提升用户的交互体验,以支持他们更好地完成工作任务。

4.4 局限性与展望

本研究也具有一定的局限性。虽然本研究开发了一套基于多维度的数字图书馆交互功能评估工具,但是要将其大范围应用还需在实践中不断验证和完善。研究的样本虽然针对数字图书馆的用户群体中的学生群体展开,然而,由于资源限制,一些用户群体被忽略,如教师群体、图书馆员等。此外,评估指标之间的关系还需进一步研究,以使评估工具进一步精炼;用户的个体特征与评估指标的关系值得探讨,以构建更为个性化的数字图书馆交互评估工具。这些也为我们今后的研究指明了方向。

参考文献

- 1 Zhang Y. Developing a holistic model for digital library evaluation [J]. Journal of the American Society for Information Science and Technology, 2010, 61(1): 88-110.
- 2 齐雪. 数字图书馆交互功能评估研究[硕士学位论文][D]. 天津: 南开大学学位论文, 2014.
- 3 李月琳, 梁娜, 齐雪. 从交互维度到交互功能: 构建数字图书馆交互评估理论模型[J]. 中国图书馆学报, 2016, 42(1): 66-82.
- 4 李月琳, 肖雪, 仝晓云. 数字图书馆中人机交互维度与用户交互绩效的关系研究[J]. 图书情报工作, 2014, 58(2): 38-46.

- 5 李月琳,肖雪,胡蝶.信息检索实验中的任务设计——真实与模拟仿真工作任务的比较研究[J].图书情报工作,2014(16):5-12.
- 6 Borlund P.Experimental components for the evaluation of interactive information retrieval systems[J].Journal of Documentation, 2000,56(1): 71-90.
- 7 闫希敏,李月琳.关键成功因素视角下的数字图书馆交互评估模型[J].图书情报工作,2016(10):24-32.
- 8 Marchionini G.Evaluation digital libraries: A longitudinal and multifaceted view[J].Library Trends,2000,49(2):304-333.
- 9 Saracevic T.Digital Library Evaluation: Toward an Evolution of Concepts[J].Library Trends,2000,49(2):350-369.
- 10 Xie H I.Users'evaluation of digital libraries(DLs): Their uses, their criteria, and their assessment. Information Processing and Management,2008 (44):1346-1373.
- 11 Tsakonas G,Mitreli A,Papachristopoulos L,et al. An exploration of the digital library evaluation literature based on an ontological representation[J].Journal of the American Society for Information Science and Technology,2013,64(9):1914-1926.
- 12 Ahmada M,Abawajy J H. Digital Library Service Quality Assessment Model[J].Procedia-Social and Behavioral Sciences, 2014,(129):571-580.
- 13 王鉴辉.数字图书馆评价体系问题研究[J].中国图书馆学报, 2004,(4):55-57.
- 14 安娜.高校数字图书馆评价指标体系研究[J].现代情报,2007,(5):97-100.
- 15 周训杰,王尊新.高校图书馆数字资源评价指标体系研究[J].现代情报,2006,(1):14-18.
- 16 王卫军.基于可用性的移动图书馆服务能力评价研究[J].情报理论与实践,2016,39(1):100-103.
- 17 何惠芬.基于用户行为的高校图书馆知识服务互动平台之构建[J].图书情报工作,2015,59:199-202.
- 18 夏立新,孙丹霞,王忠义.网络环境下数字图书馆知识服务用户满意度评价指标体系[J].图书馆杂志,2015(3):27-34.
- 19 李迎迎,王娟,郑春厚.高校图书馆数字资源服务评价指标体系构建[J].情报杂志,2014,33(3):74-81.
- 20 穆靖,刘毅,郭明明.高校数字图书馆评价指标体系研究[J].情报学,2010,28(10):1509-1512.
- 21 吴建华,王朝晖.数字图书馆评价层次分析[J].情报科学,2009, 27(8):1207-1213.
- 22 杨志和.数字图书馆发展水平的模糊综合评价[J].情报杂志, 2009(28):142-146.
- 23 刘炜,楼向英,张春景.数字图书馆评估研究[J].图书情报工作, 2007(5):21-24.
- 24 徐芳,戴炜轶.国内数字图书馆用户交互体验比较实验与分析 [J].图书馆学研究,2014,(12):18-22.
- 25 赵杨.数字图书馆移动服务交互质量控制机制研究——基于用户体验的视角[J].情报杂志,2014,33(4):184-189.
- 26 胡凯.近十年我国数字图书馆用户研究综述[J].科技情报开发与经济,2015,25(7):144-147.
- 27 黄晓斌,卢瑛.论数字图书馆用户界面的评价[J].图书馆论坛, 2005,(3):16-19.
- 28 夏春红.基于信息构建的数字图书馆评价研究[J].图书馆学刊, 2012,(3):120-123.
- 29 Chowdhury G.From digital libraries to digital preservation research: The importance of users and context[J].Journal of Documentation,2010,66(2):207-223.
- 30 Fuhr N,Tsakonas G,Aalberg T,et al.Evaluation of digital libraries[J]. International Journal on Digital Libraries,2007(8): 21-38.
- 31 梁孟华.基于用户交互的数字图书馆服务评价模型构建与实证检验[J].图书情报工作,2012,56(7):72-78.
- 32 徐芳,金小璞.基于用户体验的数字图书馆用户交互模型构建 [J].情报理论与实践,2015,38(8):115-134.
- 33 Goncalves M A,Moreira B L,Fox E A,et al.“What is a good digital library?” - A quality model for digital libraries [J].Information Processing and Management,2007,43(5):1416-1437.
- 34 Moreira B L,Goncalves M L, Alberto H F, et al. Evaluating Digital Libraries with 5SQual [C] // The 11th European Conference on Research and Advanced Technology for Digital Libraries,2007:466-470.
- 35 Blixrud L C.Measures for Electronic Use: The ARL E-Metrics Project [C] // Statistics in Practice-Measuring and Managing-IFLA Satellite Conference,2002:73-84.
- 36 Hernon P,Calvert P.E-service quality in libraries: exploring its features and dimensions [J].Library & Information Science Research,2005,27(3):377-404.
- 37 Kiran K,Diljit S.Modeling Web-based library service quality[J]. Library & Information Science Research,2012,34(3):184-196.
- 38 施国洪,岳江君,陈敬贤.我国图书馆服务质量测量量表构建及实证研究[J].中国图书馆学报,2010,(4):36-46.
- 39 郭瑞芳.基于 LibQUAL+TM 的高校图书馆移动信息服务质量探讨[J].新世纪图书馆,2013,(6):25-27.
- 40 文思敏.移动图书馆服务质量评价浅析[J].四川图书馆学报, 2013,(5):30-32.

作者单位:南开大学商学院信息资源管理系,天津,300071
收稿日期:2016年8月23日

Developing a Tool for Evaluating Interaction in Digital Libraries

Li Yuelin Zhang Xin

Abstract: This study develops a questionnaire to evaluate interaction in digital libraries based on a series of studies on the interaction between users and digital libraries and literature review. The questionnaire with 52 indicators is revised after a pilot study and totally 372 valid questionnaires has been collected via a formal survey. The degree of importance of the 52 indicators is analyzed, as well as the reliability and validity of the questionnaire. The study adopts factor analysis and identifies seven factors, such as system utility, website structure and information organization, resource accessibility, presentation of web pages, usability, task perception, and system effectiveness and efficiency. Based on the factors and the 45 indicators, the study develops an evaluation tool for the interaction in digital libraries. The study provides a valid, reliable, and operationalized evaluation tool and suggestions on how to use the tool to evaluate interaction in digital libraries.

Keywords: Digital library; Interaction in Digital Libraries; Evaluation Tool