

数字图书馆研究范畴思考

薛秋红^{1,3} 孙卫²

(1. 中国科学院文献情报中心, 北京 100190; 2. 中国科学技术信息研究所, 北京 100038;
3. 中国科学院大学图书情报与档案管理系, 北京 100190)

摘要: 数字图书馆自1994年研究、实施、应用以来, 已经在全世界的图书馆、博物馆、档案馆以及信息服务机构得到广泛的推广与使用。关于数字图书馆数字资源组织、服务成效评价、与数字出版的关系、数字资源长期保存、数字资产版权、数字素养、数字图书馆系统研究等方面的研究, 使数字图书馆的内涵更丰富, 服务成效更好, 互联网服务与图书馆服务的互补, 进而提升了图书馆资源组织和服务的能力。数字图书馆研究范畴的思考, 对数字图书馆现阶段及未来的发展和使命进行了探讨。图书馆人, 应避免以文献资料为主的研究与论证模式, 建立理论、项目、实践等分享的研究模式, 确保研究成果成为图书馆人可以借鉴、分享的果实。

关键词: 数字图书馆; 知识组织; 数字出版; 数字素养; 长期保存

中图分类号: G250.7; G254; G255

DOI: 10.3772/j.issn.1673-2286.2019.06.001

自1994年美国自然科学基金发布数字图书馆研究项目, 1996年北京召开国际图联IFLA大会, 数字图书馆从探讨性研究到广泛应用已走过了25个年头。数字图书馆是从纸质资源的数字化、资源介质化(磁带、磁盘、光盘)到数字资源网络化(服务器、远端服务器)服务的演变, 是传统图书馆资源和服务的补充, 也是网络化时代图书馆服务能力的延伸。随着计算机技术的迅猛发展, 移动互联网的普及, 数字图书馆早已突破了时间和空间的限制, 使用者可以随时随地使用资源和享受服务。在此过程中, 蓬勃发展的数字图书馆展现出强大的生命力。当今中国, 数字图书馆早已超越传统图书馆, 成为面向公众、教育、科研等各行各业最主要的信息服务手段与方法。

然而, 在数字图书馆广泛使用的今天, 依然有很多问题亟需解决^[1-3]。

(1) 数字资源的制作、采集、组织与服务依然处于种册件、整库的购置模式, 或是建立元数据并链接到资源或传统的OPAC服务模式。出现了很多与传统资源一样, 长期购买, 长期没有被利用的现实情况, 使得数字图书馆在与互联网信息服务商的博弈中, 可信赖、

有价值、更权威的资源服务处于弱势地位。在对某信息服务商研究中, 根据该服务商几千万个题名半年使用的数据统计后发现, 被使用过的题名有1 000万个, 使用2次以上的有600万个题名, 由此可见数字资源利用率不高。如何对数字资源进行组织来提高数字资源的利用率和服务的有效性值得探讨。

(2) 数字图书馆为提供服务而购买的资源与环境, 成本是固定的, 但服务成效如何? 图书馆着眼于资源数量的“性价比”, 用更少的钱买到了更多的资源(每种资源平均价格很低); 着眼于总的访问人次(点击率、检索率、下载率求和统计)。忽略了篇均利用率、有效使用率, 如何通过成效的评价, 找到改进服务效率和服务质量的地方?

(3) 数字出版已经形成相应的服务体系, 早期的数字出版以期刊库、专利库的形态出现并销售给图书馆, 目前出现了单本电子书、听书、AR结合、VR结合、流媒体等多种形态。如何找到这种数字出版资源与数字图书馆服务的关系, 并实现共赢?

(4) 检索技术、存储技术、管理与修复技术、资源组织与标引技术、用户数据挖掘技术、可视化技术、分

享与共享方法、区块链技术等,可以提升数字图书馆系统的组织和服务能力。如何面向未来,使数字图书馆系统智能化水平更高、自动化水平更高、服务的效率和质量更好?

(5) 数字资源类型从书刊报等单一数字形态变成了多媒体形态,应用从观看、视听转变到了交互,各种数字资源的应用越来越依赖于使用环境(硬件、软件、网络等)。如何从数字资源长期保存注重数据的采集和管理,转变到保证几十年甚至上百年可以应用?

(6) 数字图书馆的应用使得图书馆增加了很多数字资产,而这些数字资产的保存与利用,在现有的法律框架下也产生了不少纠纷。如何在知识产权许可下长期使用和传播这些数字资产?

(7) 美国政府国家图书馆与信息科学委员会(National Commission on Libraries and Information Science),在1974年正式提出信息素养,联合国教科文组织(United Nations Educational, Scientific and Cultural Organization)提出在学生中培养技术素养。由此可见,在科技、网络化、信息化时代,信息素养和技术素养是提高人们能力的一个重要的方向。21世纪以来,数字素养需要某些跨学科的技能,就要把信息素养、技术素养和媒体素养结合在一起,提出了信息、媒体和技术;学习和创新技能;生活和职业技能。如何使数字图书馆成为数字素养培养的基础设施?

从以上问题出发,本文从数字资源组织、服务成效评价、与数字出版的关系、数字资源长期保存、数字资产版权、数字图书馆系统、数字素养体系7个方面进行分析与思考,找出最近和将来一段时间数字图书馆领域的研究范畴。

1 数字资源组织研究

数字图书馆的信息资源主要来源于馆藏资源数字化、互联网采集和采购资源、服务等^[4-5]。资源类型有文本、图像、音频、影像、软件等,与传统图书馆的电子资源类型有很大不同。在资源组织模式上,主要包括元数据链接对象的模式、按照专题信息收集整理模式、按照购买库提供服务的模式等。这些组织模式与图书馆传统组织模式十分相似,一般是使用者通过检索的模式利用数字资源。如何提高数字内容的利用率,是数字图书馆资源组织方法值得研究与探讨的。

1.1 基于使用者需求进行资源再组织的“我的图书馆”

教育和科研类数字图书馆的用户分类清晰,阶段性目标比较明确。简单购置资源与服务的模式,对于这类用户而言,数量庞大的数字资源在使用过程中的噪声相对较大。利用资源服务系统的“我的图书馆”,依据这类用户的需求进行数字资源再组织,可以按照具体文章的分类,也可以按照期刊或者专利号进行萃取再集合。如何利用好系统中的“我的图书馆”分门别类地提供服务值得研究。

首先,要了解受众的需求,根据需求对于内容再组织。如教育类的数字图书馆,有课程大纲、课程进度、阶段练习、参考资料、实习等要求,针对这些要求和阶段进行“我的图书馆”的组织工作,引导受众直接在“我的图书馆”中利用资源,可以有效地提高购置的数字资源被有针对性地利用。

其次,要在“我的图书馆”的数字资源上进行扩展。如在“我的图书馆”中增加专利、数字书籍、数字期刊、互联网专题、多媒体培训、线上讲座等栏目,使单纯依托购置资源的组织模式,转变成满足受众需求的一站式服务的“我的图书馆”的模式。对于“我的图书馆”的数字资源还需要进行动态实时更新,资源定期组织,避免出现更新不及时的情况。

最后,在获取资源时,需与信息服务商进行合作,或用新的一个聚合系统实现“我的图书馆”功能。

1.2 基于分类法和主题词表的组织模式

期刊论文利用率更高是因为每篇文章都有中图分类号、关键词、摘要等信息,受众容易判断其是否满足需要。而对于书籍,少则几万字,多则几十万字,简单的分类或者复分,过于粗略,无法判断书籍内容是否满足需求。对每个目次内容的自动摘要能够更好地反应本目次的内容,帮助受众检索后更快地找到自己需要的数字内容。如何对书的篇章节进行分类和摘要,使检索结果更具针对性,相关方法需要研究和验证,可以使图书馆对种册件的组织模式进一步细化到篇章节的组织模式。

起源并应用于传统图书馆的分类法与主题词表是知识组织方式,最初主要用于管理资源,仅从资源管理的角度看是成功的,但是如何在数字图书馆中应用好分类、主题词,大部分图书馆没有进行研究与实践。

分类-主题词法一方面可以帮助实现自动分类与自动标引;另一方面,可以利用主题词的用代属分参族的关系扩展检索表达,利用分类进行排序,有效提高数字图书馆检索效率和质量。如爱思唯尔公司的信息服务系统,如果检索词是主题词,通过主题词的关系可以有效地帮助受众扩展检索范围,提高查全率和查准率,同时帮助受众更好地掌握主题词词汇之间的关系。遗憾的是,我国大部分图书馆都没有将分类-主题词应用于数字图书馆系统。

1.3 基于本体的知识组织模式

很多专业数字资源同质化,受众不可能完整阅读。需按照学科教育体系、科研体系等主题树的组织模式对资源基于本体的方法,重新构造知识树,并关联到实体资源。这种资源组织模式,适合于缺什么补什么和有问题快速找到答案的受众需求,是提炼精髓、快速和满足碎片化使用需求的方法之一。

本体的组织模式,大部分停留在主题词表的理论上,而主题词表只有用、代、属、分、参、族6种关系。可以认为主题词表仅仅是一个轻量本体,无法针对科研和问题进行更多关系的揭示和展示。因此,要根据不同学科的特点分门别类地建立本体。在此过程中,需要图书馆人(资源)和教育科研人员(专业)结合,专业人员解决概念树,图书馆人解决资源的分解和实体(实例)的链接。只有产学研的结合,才能逐步建立各个专业的本体知识体系,最终把研究的成果用于数字图书馆的实践。

在可视化应用上,很多图书馆混淆了本体关系可视化 and 知识图谱。在检索的噪声中,把权威的结果单独展示或提示出来,这个权威结果的展示叫知识图谱;把知识之间的相互关系展示提示出来,这个叫知识可视化,两者的本意完全不同。对于图书馆更多的应该关注知识关系的可视化,而不是知识图谱。

1.4 基于数据共享与分享的模式探索

在图书馆自动化和数字图书馆系统中,存在大量经过图书馆人整理、规范的数据。这些优质资源如何被其他信息系统所利用,共享与分享这些数据的研究及应用十分必要。

与没有经过规范化处理和系统组织的Web信息相比,图书馆作为拥有一定资源和专业化人员的机构,所拥有的数字资源是经过筛选、加工、序化和重组,具有完整性和永存性的优质资源。需要研究如何利用RDF三元组,将数据关联到互联网上,并链接到对应的实体或实例中,提高资源利用率,避免成为“僵尸数据”或“孤岛数据”。数据的共享与分享使得优质数据成为互联网重要、稳定的数据基础。

数字图书馆的资源,大部分是由高质量的出版资源构成,图书馆之间须建立合作,将沉淀在这些资源里的知识基础(如人、机构、地点、事物、事件等)进行提取并规范,建设成为基础资源,再基于这些要素对资源进行关联。维基百科就是建立在知识基础上的关联、可溯源的百科组织和服务的产品。开展知识组织和服务的准备工作是一个极具挑战性的任务。

很多机构产生的数据(如科学数据)如何被收藏、揭示与利用,是数字图书馆的一个新的研究方向。与传统的资源研究有所不同,很多科学数据是仪器设备、工具软件等产生的,再利用这些数据需考虑如何处理相应的环境,如保存不同的软件工具等对于数字图书馆是一个全新的挑战。

2 数字图书馆服务成效研究

图书馆采购的数字资源比例在上升,入馆人数没有减少,传统资源的利用率在下降。那么,是使用者不利用图书馆资源了,还是更多的在利用数字资源呢?需要对数字图书馆的成效进行分析,通过建立评价体系,找到数字资源利用提高的证据,更好地改进数字图书馆服务的效率和质量,这个研究十分必要^[6]。

2.1 资源利用率评价体系分析

类似于对传统图书馆纸质资源的评价,数字资源简单地按购买种册件的总量衡量价格,看似合理,其实不然。有60%~70%的数字内容在购买后一直是零使用,若全部折算到已用资源,则无形中抬高了每次应用的成本。应对使用的频次、使用的针对性、使用的平均成本、无效资源比例等数据进行分析并建立合理的评价体系,找到合理采购数字资源和服务的方式,最大限度地降低单篇成本,降低无效购买数量等。

2.2 资源整合体系分析

大部分图书馆通过简单的购买数字资源提供数字资源服务,通过OPAC服务,或通过购买元搜索产品对元数据进行简单的汇聚,没有将馆藏资源与购买的数字资源服务整合到一起。如果图书馆只有购买能力,没有资源组织和整合能力,当任何一个机构都可以购买数字资源并提供服务时,图书馆就失去了存在的价值。

针对科技类、教育类的数字资源,可以按照课程、项目、主题以及用户需求等,对资源进行再组织,实现资源一体化整合和关联。其主要包括元整合/过滤/排序、多媒体资源基于主题整合、知识组织与资源聚合、资源中知识基础萃取能力等方面的分析与评价相关的研究。

2.3 数字图书馆服务方式与服务效果评价

大多数图书馆仅提供信息检索与链接服务,或者IP范围内信息服务商的服务,忽视了栏目分类、主题服务和专题服务等针对用户习惯和需求的服务方式的设计。如何在数字图书馆系统中,根据不同的用户习惯提供相应的服务方式,根据不同的用户需求,提供合适的资源组织,改进数字图书馆获取资源的简单服务模式;通过对数字图书馆服务方式和效果的评价,对问题、检索、参考咨询整合式的服务模式的研究,文献传递、检索服务、参考咨询服务组合式的关联的效率和质量的评价,是提升数字图书馆服务效率和质量的重要研究方向^[7]。

2.4 用户对数字图书馆的满意度评价

用户的满意度评价对改进数字图书馆的使用效率和服务质量非常有意义。如检索结果的召回率、准确率、输出形式、响应时间、时效性、用户依赖性、虚拟参考咨询是否解决用户问题等^[8],这些研究对改进数字图书馆的服务模式十分有价值。

大部分图书馆针对馆藏资源进行简单的入馆培训,特别是停留在新生对于大学生活还一无所知阶段的程序化培训上。对于进一步扩大图书馆用户使用数字图书馆处于被动的状态,研究数字图书馆对于用户的作用,是改进被动服务为主动提供有针对性服务的有的放矢的方法之一^[9]。

2.5 对数字图书馆系统的评价研究

目前的数字图书馆系统主要是多种分布式技术产品的整合,在整合、标准化等方面缺乏研究和评价。特别是移动终端使用数字图书馆,大部分停留在检索、预约、续借、参考咨询阶段,资源服务的提供能力有限。很多图书馆的技术部门停留在已有系统的安装调试维护上,利用开源软件开发与构建数字图书馆的技术能力需要研究和评价。特别地,对于未来技术如何用于数字图书馆建设和服务的研究与评价,是保障数字图书馆可以持续提供高效和高质量服务的关键因素之一。

2.6 对图书馆馆员在数字图书馆建设与服务能力方面的评价研究

图书馆属于人力资源、财政资源有限的机构之一。如何提高人力资源和财政资源在数字图书馆建设和服务中的效率和质量是一个很重要的问题。数字图书馆与传统图书馆是互补的关系,那么在建设数字图书馆时,首先要考虑数字图书馆使用的效率和质量,馆员需要充分了解本机构的需求、本馆技术装备适应的范围、数字图书馆资源的试用与评价、不同服务商的能力评估、不同服务商的资源覆盖与交叉的评价以及更换服务商后服务水平和服务能力风险的评估等。可见,在数字图书馆建设与服务过程中,对于馆员的要求是截然不同的,有必要建立一套针对数字图书馆建设和服务的馆员能力与财务支撑能力的评价体系。

3 数字图书馆与出版商的关系研究

3.1 数字内容二次传播与出版商的关系

目前数字图书馆以购买出版商和信息服务商的信息服务为主。这种模式的一个很大弊端,就是只能按照卖方构建的体系购买相对集中的内容,实际上很多内容购买以后就是零利用率。传统图书馆不会购买大量的期刊,只会针对读者的需求购买,但是到数据库服务时,就变成了付费买了很多无用的资源。那么研究数字图书馆与出版商的内容和服务联动机制也是一个方向。如亚马逊与图书馆达成一种图书馆电子书利用协议,即Kindle格式的电子书,图书馆付一次费可以下载26个副本进行使用等。国内繁多的电子书系统(如喜马拉雅、咪咕

等), 如何与图书馆现有系统结合, 扩展资源组成结构, 构成新的资源发现系统, 是一个重要的研究方向。

数字图书馆涉及到数字资源的二次传播, 与传统图书馆只有一个副本提供服务后还回是不同的资源利用模式。研究与出版商之间达成协议来避免图书馆购买低利用率资源, 需要时二次传播又面临出版商诉诸法律的尴尬局面。加州大学图书馆就与出版商达成了一个馆购买, 多个馆使用的协议; 亚马逊的数字图书也与图书馆达成协议, 每个付费可以进行26次下载, 第27次下载时, 图书馆需要第二次付费。研究出版商与图书馆都能接受的协议, 在我国还是一个空白。

涉及馆内有纸质副本, 传递相应的数字资源也存在一定的风险, 这也值得研究。如NSTL购置了外文期刊, 万方数据揭示这些期刊, 本意是提高这些期刊的再利用, 而外国出版商以万方数据没有购买期刊不能分享元数据为由, 与NSTL和万方数据进行磋商。由此可见, 传统文献传递变成数字形态的传递也需要进行研究。

3.2 机构库与出版商之间的关系

目前, 图书馆在大量构建机构库, 机构库的内容除了本机构的很多私有的内容(PPT、资料等)以外, 还收集到很多非本机构产生的公开的出版内容。这些内容没有得到出版商和作者的授权, 在私人收藏与研究时, 作为参考资料不涉及知识产权和互联网传播权的问题。当图书馆构建机构库时, 对于出版内容进行集藏、提供检索服务、二次传播就涉及知识产权问题。如图书馆扫描图书, 放在机构库, 提供给本机构人员下载使用, 由于未经作者和出版机构授权, 涉及数字化传播侵权。又如, 某图书馆未经出版机构和作者授权, 通过数字化传递, 把数字化后的内容传播给了需要者, 被作者和出版机构获得证据诉诸法律。所以, 如何使用与传播这类资源需要进行研究。

3.3 图书馆的能力与出版商之间的关系

图书馆在建立出版物的分类、主题词、索引、导读方面具备较好的能力, 可以借此帮助出版商更好地进行出版和提供更准确的数据, 国外很多出版商在出版前与图书馆进行了合作。

中国的出版在学术出版领域具有大量的出版物, 已

经在制定标准, 规范学术出版。但如何查证学术出版的参考资料? 如何做学术出版的索引? 在国外图书馆与出版机构合作建立高质量的出版有很多实例, 是一个值得研究的方向。

3.4 开放存取与出版商之间的关系

在2004年中国图书馆进行开放存取的研究以来, 并没有在商业价值链的角度很好地研究如何建立合作共赢的开放存取体系, 使其具有中国特色, 同时又是与出版合作的楷模^[10]。如当购买的资源有本机构的资源内容时, 如何减少购买的费用? 再如, 开放存取有时间期限, 期限未到时, 如何利用? 只有和出版商建立多赢的模式, 开放存取才真正具有生命力。

4 数字资源长期保存的研究

保存人类记忆遗产是图书馆的重要职能。因为有了图书馆, 人类社会实践所取得的经验、文化、知识才得以系统地保存并流传下来, 成为人类宝贵的文化遗产和精神财富。在图书馆里, 可以看到3 500年前的甲骨文, 1 500年前的宣纸, 100多年前的电影胶片等载体类型的文献资料, 这些资料不仅被长期保存, 而且至今可以供人们阅读与使用。但是, 很多电子资源(唱片、CD、VCD等)、数字资源和网络资源受到诸多因素的限制, 保存时间只有几十年, 尽管载体完好, 但不能再被人们阅读与使用, 失去了再利用价值。若继续保存这些资源, 一是占用图书馆的空间, 二是需要资金持续投入购置设备和改善存储环境。目前国内的数字资源长期保存技术停留在内容的收集与管理上, 在迁移、转换与仿真等方向缺乏研究。

4.1 迁移技术研究

早期的数字资源长期保存研究, 主要是基于OAIS模型的获取、存储与管理。而从不同的介质上进行转移数字资源的过程, 称之为迁移。软盘、磁带、光盘、磁盘、半导体等存储介质, 都存在被淘汰的风险, 需要进行新介质的迁移, 而在应用介质迁移技术时, 还需对其成本和可修复性等进行考虑。如主要存储在软盘、磁带上的内容需要迁移到光盘、磁盘或者新的介质上, 需要研究不丢失信息内容、如何比对和修复等技术。迁移技术研

究需要图书馆重视,软盘、光盘、磁带、录音录像带已经出现介质失效或者环境不存在而无法使用的高风险。

4.2 转换技术研究

不同的数字资源对应的文件格式不同,有的格式是标准的,有的是非标准的。对于淘汰的格式标准或者非标准的数字对象,就需要将原有格式转换成新格式或是标准格式,以便再利用。在研究文件格式转换技术时,需对其再应用时的标准性和转换后的保真性等进行思考。由于数字资源已经从传统的图书、期刊、报纸的数字化转变成混合模式的多媒体和人机交互式的资源对象,所以要思考对于这类资源的转换技术的研究与应用。如FLASH技术的流式格式如何转换到H5技术的内容就是一个典型的格式转换技术的挑战。

4.3 仿真技术研究

在数字出版刚起步时,相当一部分数字出版物是借助互联网传播的,所以,作为网络的节点,计算机无疑是数字阅读最早的终端设备。随着技术不断提高,计算机由曾经的台式机发展到笔记本、平板电脑、移动终端。越来越多的数字资源与阅读的设备、软件等环境相关。保存与维护这些设备和软件是一笔巨大的投入,在图书馆长期保存与维护这些设备和软件已不大可能。应对的策略就是仿真这些设备和软件的环境,这是一个全新的研究与挑战。在计算机上依然可以仿真各个版本的DOS、Windows等,所以需探讨如何利用仿真环境来替代昂贵的设备和软件的保存方式。在移动设备上也存在仿真各种环境的需求,需研究与验证仿真环境的技术,并对于各种相应的数字资源在仿真环境下的可利用性进行检验。

4.4 专用软件环境研究

由于数字图书馆涉及科学数据的保存与利用,而这些科学数据大部分不同于传统数字图书馆的数字资源,很多是由仪器设备与专用的软件工具产生的,由于仪器设备和软件工具的专用性,简单的仿真是无法保证这些数据的再利用的,那么就需要研究哪些专用软件因为哪些特殊用途而需要保存并进行维护,使科学数据可以再利用。

此类问题在数字素养的培训、训练中也会遇到。所以,数字资源与使用环境的研究成为一个新的研究方向。

5 数字资产版权研究

传统图书馆都是以完整的种册件形式和少量的复制形式提供资源服务的,鲜有著作权纠纷。在数字图书馆时代,未经作者和出版机构授权的数字化加工、使用未经查实的数字资源提供服务、直接利用数字化资源进行二次传播等过程中都或多或少地与作者、出版机构发生各种法律纠纷。

5.1 数字资产版权类型的研究

很多图书馆简单地利用著作权、互联网传播权、《图书馆法》中的某些条款,试图规避风险。一旦发生纠纷,则采用简单关断服务的做法,直接影响了用户利用数字资源的权利。面对传统资源数字化,涉及表现形态的变化与互联网传播授权两个风险点,需要研究数字资产版权类型的变化,如二次传播权、传统资源与数字资源内容一致时的同权、终端用户获得并使用数字资源的权利、存储数字资源或收藏对应纸质资源的权利等。需要研究在现有权利范围规避侵权的方式方法,避免图书馆陷入侵权,成为被告的尴尬局面。如国家图书馆、中国科学技术信息研究所、中国科学院文献情报中心等机构尝试与出版机构达成协议的模式来解决数字资产的权利风险。

5.2 数字资产损失的风险

对于传统资源,购置后是放置在图书馆,只要图书馆在,资源就可以阅读与使用。早期的电子资源,也是有磁带、光盘等介质的,也具有很好的使用权。但是,现在的信息服务商提供的服务,数字资源是在信息服务商的服务系统中,数字图书馆的服务在各个图书馆里面,一旦图书馆不购买服务,已经购买过的数字资源服务也就停止了。国家图书馆、中国科学院文献情报中心等机构,对于购买过的数字资源保留服务权利和方法进行了很多尝试,值得借鉴和研究。但是,大部分图书馆并不具备这样对等谈判的优势,在此背景下,图书馆是否可以在不购买服务或服务商停止服务时,获得资

源,继续为用户服务,值得思考。

5.3 数字资产技术保护的解锁

大部分信息服务商和出版商会利用技术手段保护各自的商业化产品,面对很多采用技术保护的数字资源,在出版商和信息服务商失权后(如出版商的邻接权为10年)面临解除技术保护的难题,如何解决这个问题,使得数字资源可以在解除技术保护后继续使用值得研究并提出对策。

6 数字素养体系研究

国际图联在20世纪提出信息素养,后来出现技术素养,都是在互联网信息爆炸时代、技术工具大量涌现时代提出的对人的基本能力的要求。数字素养是信息素养、技术素养^[11]、科学素养、逻辑与计算等综合素养在21世纪新的提法,从儿童开始进行系列化的数字素养教育与培训。

目前我国把信息素养、技术素养当作应试或选修课程,多数人并没有真正提高自己的信息素养和技术素养,一些传统图书馆甚至把信息素养的教育片面到检索方法的使用上。

人类在进入网络时代、知识社会后,需要从儿童抓起,提倡数字素养教育。数字素养主要包括7方面的内容。①硬件和软件的基础知识。②信息和数据素养。对信息内容的浏览搜索过滤、评价、管理等。③沟通与协作。通过数字技术进行互动、共享、参与、协作,还包括网络礼仪、管理数字身份等。④数字内容创作。开发数字内容、整合重新构建数字内容、版权和许可、编程等。⑤安全。保护设备、保护个人数据和隐私、保护健康和福祉、保护环境等。⑥解决问题。解决技术问题、确定需求和技术响应、创造性地使用数字技术、确定数字能力差距、计算思维等。⑦与职业相关的能力。

数字图书馆平台需要担负起数字素养教育培训的工作^[12],体系需要进一步拓宽,以适应很快出现的数字素养教育体系的需要^[13]。提升数字素养也会对数字图书馆的建设和服务产生影响。

7 数字图书馆系统研究

数字图书馆的相关技术研究围绕数字图书馆建设

和服务效率与质量的提高,不断应用与更新先进技术的迭代过程,利用技术预测来评估未来的技术体系和架构^[14-15],数字图书馆系统需要不断研究和跟进。

7.1 检索技术

检索是数字图书馆资源发现的基本功能,目前检索的效率和质量都还有很多值得改进的地方:①改进检索效率和质量,研究排序、过滤、多面浏览;②检索系统的整合,研究本地检索与互联网检索系统的联动、过滤、结果展示;③研究检索内容链接的有效性技术,很多检索结果的链接是无效的,使用者点击造成失败,发现并过滤掉非有效链接是避免检索噪声的一个技术研究方向;④可视化技术的研究,使得数字资源的知识组织、主题组织进行可视化的展示,便于使用者按图索骥找到数字资源。

7.2 知识组织与资源关联技术

资源的知识体系架构与用户需求体系架构的匹配化研究,使得粗略的检索、平铺、大噪声等向高效、准确、权威、参考关联服务模式转变。①研究对于种册件数字资源的再标引、抽取、自动摘要等技术,使得资源服务满足碎片化的需求;②研究从资源中萃取知识基础的理论和标准,建立技术体系,从资源中提取知识基础,为资源的复用、关联、参考关联等建立分布式的体系架构;③研究如何从轻量本体、复杂本体、主题图等模式与资源链接的技术,并与已有的检索系统、可视化系统结合,展现知识结构和关联资源。

7.3 数据挖掘技术

图书馆的借阅数据、网站日志、数字图书馆使用日志,都是读者使用图书馆和数字图书馆的有关信息。如何从这些信息中,挖掘出读者需求与图书馆和数字图书馆的关系、一类用户在时间需求阶段一致性的关系、科研用户在申请课题与研究过程中与数字图书馆的关系等,都是需要针对需求、阅读关联性(用户理解、内容理解、内容与用户的匹配)等进行挖掘,找到提高服务针对性的要素,改善服务的效率和质量。

图书馆特别是数字图书馆有大量的二次文献和一次文献,对于主题、分类的精准性以及链接资源的权威

性进行挖掘,对改进检索结果的展示非常重要。

对用户需求和资源挖掘后进行匹配的研究,数字图书馆建设和服务利用这些研究建立后台的计算模型,能更好地改进数字图书馆系统的服务。

7.4 区块链技术

区块链技术主要基于去中心、加密、业务证据不可修改性而备受推崇。对于数字图书馆,自建资源和购置资源在提供数字图书馆服务时,会出现总分馆模式、联合购买使用模式,那么去中心和业务证据不可修改方向可能是一个应用的机会,对于加密需要与信息服务商进行合作解决。研究区块链技术如何在分布式数字图书馆中使用,平衡资源分享、共享、保护知识产权的目的。需要从现实的数字图书馆系统着手研究区块链技术,并进行实验。避免空洞的、不可操作或者业务模式不存在的区块链技术研究。

7.5 资源长期保存技术

数字资源长期保存资源迁移、转换、仿真技术、技术保护解除等的研究^[16-17]。像古籍保护与修复技术一样,研究长期保存中的修复技术,是保证数字资源在介质有效、格式有效、环境有效前提下的一个重要的数字资源修复能力的体现。研究科学数据保存中,除了已有的揭示、保存以外,如何保存科学数据的运行环境也是一个重要课题,否则大量的科学数据随工具的变化或消失成为不可用的科学数据。特别是要对科学数据类型、产生环境、复用科学数据的环境进行研究。由于很多图书馆开始租赁云环境,软件在不同云平台的运行、迁移也是一个值得研究的方向,确保在云迁移的过程中,数字图书馆系统的服务能力、可靠性、性能的不降低。运行数据的安全性、去隐私性、分享日志文件等方面是使用云平台后的挑战。

7.6 建立数字素养培训平台

对于数字素养的培训,需要建立课程体系(MOOC),对公众进行教育。对于学习课程以后,建立训练实操的过程和抽取数字图书馆内容在检索、溯源、判断真伪、汇总、摘要、统计、评价等提供数字素养训练是任务。在数字图书馆平台建立一些软件工具,利用数字图书馆

的资源集、使用数据集与工具等,培训公众利用这些工具的能力。研究数字素养教育中需要的相关工具和技术,将数字图书馆平台同时作为数字素养学习^[18]、训练的平台^[19]。

8 结论

综上所述,数字图书馆作为图书馆资源与服务的补充,还有很多值得研究与实践的地方,目的是提高数字图书馆建设和服务的效率与质量,更好地为教育、科研提供满足用户需求的服务系统。

图书馆人,应避免文献资料研究为主模式,应该加强理论、项目推进的可分享的研究模式。问题、论证、结论、评价是一个论文所必须的4个要素,数字图书馆的研究应该是以需求问题、假定证明两个为导向的研究与实践。确保研究成果成为图书馆人可以借鉴、分享的果实,真正成为研究、验证、应用的典范之一。

参考文献

- [1] LYNCK J. Library Tech Trends for 2019 [EB/OL]. [2019-05-20]. <http://www.techsoupforlibraries.org/blog/library-tech-trends-for-2019>.
- [2] PEACHEY J. How people in the UK and Ireland use public libraries and what they think of them [EB/OL]. [2019-05-15]. https://d1ssu070pg2v9i.cloudfront.net/pex/carnegie_uk_trust/2017/04/LOW-RES-2867-Libraries-Main-Report-Complete-1.pdf.
- [3] PEACHEY J. The future of public libraries across the UK and Ireland [EB/OL]. [2019-05-15]. https://d1ssu070pg2v9i.cloudfront.net/pex/carnegie_uk_trust/2017/04/Policy-Report-Shining-a-Light.pdf.
- [4] ARMS C, FLEISCHHAUER C, MURRAY K. Sustainability of digital formats planning for Library of Congress collections: external dependencies [EB/OL]. [2019-05-15]. <https://www.loc.gov/preservation/digital/formats/sustain/sustain.shtml>.
- [5] Library of Congress. Library of Congress Recommended Formats Statement [EB/OL]. [2019-05-24]. <https://www.loc.gov/preservation/resources/rfs/RFS%202017-2018.pdf>.
- [6] BERTOT J C. Assessing digital library services: Approaches, issues, and considerations [C] //International Symposium on Digital Libraries and Knowledge Communities in Networked

- Information Society (DLKC' 04), 2004.
- [7] AHMAD M, ABAWAJY J H. Digital library service quality assessment model [J]. Procedia-Social and Behavioral Sciences, 2014, 129: 571-580.
- [8] RAFI M, ZENG J M, AHMAD K. Evaluating the impact of digital library database resources on the productivity of academic research [J]. Information Discovery and Delivery, 2019, 47 (1): 42-52.
- [9] VRANA R. The perspective of use of digital libraries in era of e-learning [C] //2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). IEEE, 2017: 926-931.
- [10] GUÉDON J C, KRAMER B, LAAKSO M, et al. Future of scholarly publishing and scholarly communication: Report of the Expert Group to the European Commission [EB/OL]. [2019-05-24]. <https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1098&context=scholcom>.
- [11] ANYIM W O. Assessment of ICT literacy skills of digital library users and Staff in Salem University Lokoja, Kogi [J/OL]. Library Philosophy and Practice, 2018 [2019-05-15]. https://www.researchgate.net/publication/325052795_Assessment_of_ICT_Literacy_Skills_of_Digital_Library_Users_and_Staff_in_Salem_University_Lokoja_Kogi.
- [12] UNESCO Institute for Statistics. A Global Framework to Measure Digital Literacy [EB/OL]. [2019-05-27]. <http://uis.unesco.org/en/blog/global-framework-measure-digital-literacy>.
- [13] Bloom-Carroll Local School District. Technology Plan 2017—2020 [EB/OL]. [2019-05-27]. <https://www.bloom-carroll.k12.oh.us/Downloads/District%20Tech%20Plan%202017-2020%20w%20TOC%20revised%2008262017.pdf>.
- [14] Gaetano Cascini. Introduction to Technology Forecasting [EB/OL]. [2019-05-15]. <http://www.oiepec.eu/wp-content/uploads/2017/07/Introduction-to-Technology-Forecasting.pdf>.
- [15] DECLERCQ J F, CALLEWAERT R, VERMAUT F. Flemish Public Libraries – Digital Library Systems Architecture Study [EB/OL]. [2019-05-15]. <http://istoiresservices.be/wp-content/uploads/2018/11/2013-11-27-DigitalLibrarySystemArchitecture.pdf>.
- [16] Digital Preservation Methods and Technology In Digital Era [EB/OL]. [2019-05-14]. https://www.academia.edu/13486465/Digital_Preservation_Methods_and_Technology_In_Digital_Era.
- [17] Protecting & Preserving Long-Term Digital Information for IT Professionals and Practitioners [EB/OL]. [2019-05-14]. https://www.statearchivists.org/files/1714/8589/3400/Protecting_and_Preserving_Long-Term_Digital_Information_COSA_Jan17_v07.pdf.
- [18] Information and Digital Literacy Tutorials [EB/OL]. [2019-05-14]. <https://www.sheffield.ac.uk/library/idlt/preentry>.
- [19] A Global Framework of Reference on Digital Literacy Skills for Indicator 4.4.2 [EB/OL]. [2019-05-14]. <http://uis.unesco.org/sites/default/files/documents/ip51-global-framework-reference-digital-literacy-skills-2018-en.pdf>.

作者简介

薛秋红, 女, 1990年生, 博士研究生, 研究方向: 知识组织、智能信息处理, E-mail: xueqiuHong@mail.las.ac.cn。

孙卫, 男, 1957年生, 计算机高级工程师, 中国科学技术信息研究所客座研究员, 研究方向: 数字图书馆、数字出版、知识组织, E-mail: sunwei@istic.ac.cn。

Thinking on the Research Field of Digital Library

XUE QiuHong^{1,3} SUN Wei²

(1. National Science Library, Chinese Academy of Sciences, Beijing 100190, China;

2. Institute of Scientific and Technical Information of China, Beijing 100038, China;

3. Department of Library, Information and Archives Management, University of Chinese Academy of Science, Beijing 100190, China)

Abstract: Since the research, implementation, and application of digital libraries in 1994, it has been widely used in libraries, museums, archives, and information service organizations around the world. Digital library research in digital resource organization, service effectiveness evaluation, relationship with digital publishing, development of digital library system, long-term preservation of digital resources, copyright of digital assets, improvement of digital literacy and other aspects can make the connotation of digital library richer and service effect better. In the future for a long time, library services need to complement the services of the Internet to enhance the library's capabilities. The thinking of the research scope of digital libraries is the discussion of the development and mission of the digital library at this stage and in the future. Librarians should avoid research and argumentation models based on literature, and it is necessary to establish research models that can be shared by theory, projects, practices, etc., to ensure that research results can be the fruits that library people can learn from and share.

Keywords: Digital Library; Knowledge Organization; Digital Publishing; Digital Literacy; Long-Term Preservation

(收稿日期: 2019-05-08)