

# 大数据时代数字图书馆面临的机遇和挑战<sup>\*</sup>

苏新宁

**摘 要** 大数据时代,网络学术资源建设快速发展,大有取代数字图书馆之势。在这样一种背景下,数字图书馆建设要有大数据思维:在资源建设方面,要拓展资源范畴,增加资源整合的广度,加强资源组织加工的深度;在技术应用方面,重视语义技术,加强聚类技术的应用,广泛应用数据分析技术,提升检索技术与水平;在服务方面,丰富数字图书馆服务产品,服务模式由被动、等待、共性化转向积极主动、自动推送和个性化。未来数字图书馆要跳出传统资源建设的框架,树立全方位服务的理念,提供全资源服务,拓展服务方式和传播形式,数字图书馆馆员应该立志成为数据分析家。参考文献 18。

**关键词** 大数据 数字图书馆 资源建设 服务理念 服务模式

**分类号** G250

## Opportunities and Challenges Faced by Digital Libraries in the Era of Big Data

SU Xinning

### ABSTRACT

With the rise of academic resource database vendors in the big-data era, several important vendors in China, such as CNKI, Wanfang, Weipu and Superstar database as well as some online resource vendors, such as Baidu, are developing rapidly in the resource construction in respect to the academia, education and culture. These resource construction vendors are continuously expanding their information resources and at the same time diversifying the forms of service. It will cause great threat to the survival of digital libraries. The issues such as how the digital library survives and develops, whether it will be replaced etc, must be paid great attention to. Therefore, the original concept about resources, organizational resources, technology, services and functions of digital libraries should be changed, and the infrastructure of digital libraries should be reconstructed. In this way the current challenges might be turned into opportunities for the rise of digital libraries.

The article reviews several related research works and summarizes some important researches and their main achievements in the big data era of digital libraries. Learning from the resource vendors, we propose to use big data thinking to construct digital libraries in the era of big data. The major points of this paper are described as follows:

<sup>\*</sup> 本文系国家自然科学基金面上项目“面向知识服务的知识组织模式与应用研究”(编号:71273126)的研究成果之一。(This article is an outcome of the project “Knowledge service-oriented knowledge organization model and application research” (No. 71273126) supported by National Natural Science Foundation of China.)

通信作者:苏新宁,Email:xnsu@nju.edu.cn,ORCID:0000-0003-1632-8716(Correspondence should be addressed to SU Xinning,Email:xnsu@nju.edu.cn,ORCID:0000-0003-1632-8716)

In the aspect of the development of a digital library, we propose to extend the resources of digital libraries by including not only the information in literature but also the information from governments, network with broad social concerns; We also suggest to expand the scope of resource integration in digital libraries, which aims at not only integrating resources within the library, but also integrating the information resources from social point of view and from the network platform, especially integrating the resource useful in solving complex problems; furthermore, to improve the depth of process for digital library resource organization, it is necessary to pay more attention to the deconstruction and reconstruction of knowledge during resource processing.

In the aspect of technology application for digital libraries, we propose to adopt technology relative to semantic web application, and to build a framework able to link different types of semantic relationships from heterogeneous resources, such as integrating the semantic relationship implied in existing thesaurus and dictionaries into digital resource description, improving the current clustering algorithms to make the data more fit for specific problems and able to detect the patterns in the dataset. We discuss the widely used technologies for data analysis, because in the future, the digital library will live on the resources and depend on these technologies to provide high-quality data products. We emphasize the importance of grid search method about cross-platform and cross-resources seamless search for future digital library, which can provide more powerful semantic retrieval results to enhance the performance and quality of information retrieval.

In the aspect of digital library services, we emphasize the expansion of service products of the digital library. Apart from providing the literature information and related information on the internet, a digital library should be able to provide the think-tank products, such as the product of recycled resources, the product of information analysis, and the video product of science, education and culture resources. Moreover, the service mode of digital libraries should change from passive and general services to active, automatic push and personalized service mode.

When considering the position of strategic development of digital library, we believe it is essential to jump out of the conventional infrastructure of resources in digital libraries. The digital libraries should provide omnibus information service based on the full resource services and further extend the types of service and the ways of transmission. Digital libraries also need to shift some part of their focus to provide high-end service, such as to support decision-making or to tackle key research projects. Finally, we suggest new standard of requirements of being a librarian, which requires every librarian should aspire to become a data analyst.

The contribution of this paper is that it proposes the orientation of developing the digital library in the future and discusses the related potential problems. 18 refs.

## KEY WORDS

Big data. Digital libraries. Resource development. Service ideas. Service mode.

## 0 引言

2012年12月,马云和王健林关于“未来若

千年内,电子商务能否取代传统实体零售”的辩论,虽然还没有最终结论,但2014年“双11”那一天,仅仅阿里巴巴旗下天猫淘宝的销售额就超过500亿元。这促使我们许多人开始重新思

考电商在未来若干年能否超过传统商业。由此而来,也将引发我们的焦虑:大数据时代,传统的数字图书馆是否会被新兴的网络资源建设商所取代。提出这个问题,不是危言耸听,也不是蛊惑愚众,这是一个现实所在。

目前,学术资源建设商发展迅猛。例如,中国最大的学术资源建设商,中国知网(CNKI)资源建设已由中国期刊论文逐步拓展到国内外期刊论文、会议论文、学位论文、重要报纸文章、专利、标准等,在图书资源方面,也收录大量图书,不仅提供年鉴、工具书的查询服务,最近还建立了教辅平台<sup>[1]</sup>。另一重要学术资源服务平台——万方数据知识服务平台<sup>[2]</sup>,也提供了国内期刊论文、学位论文、会议论文、专利、标准、地方志、法规文献、科技成果、图书、行业机构、专家学者等学术资源。还有其他的资源建设商,如维普期刊服务平台等在学术资源建设方面都取得了许多成绩。目前,这些资源建设商所拥有的非图书资源是国内绝大多数图书馆都无法比拟的。我国大多数高校图书馆已经离不开这些资源建设商提供的学术资源服务了。

作为以图书提供服务为主的“超星数字图书馆”,由北京世纪超星信息技术发展有限责任公司建设,该数字图书馆包括文学、经济、计算机等五十余大类,拥有数百万册电子图书,500多万篇学术论文,全文总量超13亿页<sup>[3]</sup>,已经达到一个大型图书馆的藏书规模,假以时日,将会超过目前我国绝大多数图书馆的藏书。另外,该数字图书馆还收藏了近20万集<sup>[3]</sup>的学术视频,这些收藏对我国传统数字图书馆形成了很大的冲击。

除了以上以文献资源建设为主的学术资源建设商,一些网络资源服务商也开始涉足文献资源。例如,百度已进入学术资源领域<sup>[4]</sup>,除了百度文库、百度百科等学术资源,还提供数十万种图书供阅读,2014年上线的百度学术搜索更是提供了传统数字图书馆不能提供的学术信息,百度学术搜索是提供海量中英文文献检索的学术资源搜索平台,可以一站式检索到收费

和免费的学术文献,是学术研究的好帮手。据测算,目前百度拥有的信息总量超过1 000PB,相当于国家图书馆藏书数字化后数据量的5万倍,每天增加的数据量10TB,相当于半个国家图书馆的藏书数字化后的数据量<sup>[5]</sup>。当然,绝对用百度的数据量和国家图书馆的藏书数据相比较也许不一定合适,因为在数据质量和规范上,国家图书馆拥有更大的优势,而且现在国家图书馆也收藏了许多数字资源、图像视频资料以及互联网上的信息。本文在这里的数据量上的对比,主要为了引起图书馆人的重视,并对图书馆界产生警示。

由此可见,这些网络资源建设商在学术、教育、文化资源的建设上发展十分迅速,不论是资源的规模还是资源的种类,已经是图书馆所建的数字图书馆所无法比拟的。这些资源建设商对资源的开发还在深入和拓展,服务的形式将会更加多样化,如此发展下去,数字图书馆将面临很大的威胁,未来发展会受到严峻的挑战。数字图书馆将如何生存,如何发展,是否会被取代?这是图书馆必须要重视的问题。

当然,这种挑战也可能成为图书馆的机遇。图书馆必须借助大数据时代这一机遇,转变原有的对资源、资源组织、技术、服务以及职能等方面的认识,重新架构数字图书馆,把挑战变为再一次腾飞的机遇。这即是笔者撰写此文的目的所在。

## 1 相关研究与启迪

大数据时代,人们的思维发生了根本转变。对于数字图书馆而言,也应当接受大数据的思维。为了得到相关研究的启示,笔者检索了国内(CNKI)和国外(EBSCO)全文数据库,同时涉及大数据和数字图书馆主题的文章分别检索出40余篇和20余篇,真正将大数据与数字图书馆密切关联的文章只有10余篇,其中国外文献只有寥寥几篇。但这些文章对我们思考数字图书馆的未来具有一定启示作用。

## 1.1 国内外相关论述

大数据时代,人们在不断探求大数据与数字图书馆的融合点,国外学者在理念、技术和应用方面做了许多研究。美国加州大学伯克利分校图书馆的 Huwe<sup>[6]</sup>在《构建数字图书馆》一文中指出:大数据与图书馆是天作之合,图书馆员对用户进行研究的项目尤其适合大数据,希望能够加强数字图书馆对用户行为信息的采集。加州大学尔湾分校的 Renaud、麻省理工学院的 Britton 等人<sup>[7]</sup>借助大数据技术,对使用大学数字图书馆用户的行为进行深度挖掘,并帮助学校对学生阅读行为及相关信息进行关联分析。在技术方面,美国加州大学洛杉矶分校的 Borgman 等人<sup>[8]</sup>利用嵌入式技术,在数字图书馆系统中嵌入传感器,采集有关数据,为研究人员开展研究提供数据,为数字图书馆增添了新功能。还有许多学者已经开始关注数字图书馆对大数据中学术信息的采集、处理、关联<sup>[9]</sup>,使数字图书馆更好地融于互联网,更好地利用大数据。

在国内研究方面也有许多相关成果和理念。如,曾建勋在《数字图书馆论坛》的“大数据与数字图书馆”专辑的卷首语中指出<sup>[10]</sup>:大数据必将促进数字图书馆数据管理、数据分析、数据使用及数据服务的深层次变革,同时也对图书馆员的职责和意识提出了更高的要求。武汉大学陈传夫等人<sup>[11]</sup>分析了大数据环境下数字图书馆面临的问题,并从转变观念、发展知识服务、完善财政投入机制、提高图书馆员素质等方面提出了有关建议与对策。上海图书馆的刘炜等人<sup>[12]</sup>强调在大数据时代数字图书馆应关注大数据与关联数据,展望“大”关联数据和关联大数据两种不同的关联数据在数字图书馆中的应用前景。兰州商学院的陈臣<sup>[13]</sup>从技术角度探讨如何建立高效的、满足用户个性化需求的数字图书馆搜索引擎。在大数据技术应用于数字图书馆的其他方面也有不少研究,陈茫等人<sup>[14]</sup>研究了大数据环境下数字图书馆的移动服务技术,陈臣<sup>[15]</sup>、李白杨等<sup>[16]</sup>浅议了数字图书馆中

的大数据存储问题,郭春霞<sup>[17]</sup>涉猎了大数据环境下非结构化数据的融合问题,王宇鸽等人<sup>[18]</sup>探讨了数字图书馆中大数据技术应用架构问题,等等。

以上研究说明,学界和图书馆界已经意识到大数据对数字图书馆带来的挑战和机遇,也正在进行数字图书馆和大数据融合的研究与实践。为了使高校图书馆和公共图书馆的数字图书馆在大数据时代得到更好的发展,并在人们的文化生活和科学研究中发挥更大作用,我们还应当深入思考,转变观念,提升竞争力。

## 1.2 启示与思考

学术资源建设商在大数据时代得到迅猛发展,其涉足领域不再限于网络资源和期刊论文,已包括所有类型的文献资源,资源的组织也不再是简单的文献组织,已经将文献间的关联关系(引用、同被引、引用耦合等)、作者与文献的关系、文献的使用信息、学者之间的关系等都建立联系,这些都是传统数字图书馆没有做到的。随着这些学术资源建设商的资源不断丰富,极有可能超越传统数字图书馆,有取代高校和公共图书馆建立的数字图书馆之势。

因此,图书馆界应当引起重视,把来自资源建设商的威胁当作又一次勃发的机遇。图书馆可以汲取资源建设商的经验,从资源采集的范围着手,扩大数字图书馆的资源范畴,加强各类资源的融合,探索各类资源间的相互关联,并将它们有机组织在一起。在资源价值的发现方面,做到充分发挥数据的作用,对所采集的数据进行深度挖掘,找到其潜在规律,对相关数据进行深入分析,发现其内在价值。从服务角度,图书馆也必须学习资源建设商的一些经验,提供多种多样的服务形式,改变传统数字图书馆被动式(用户要什么给什么)、等待式(用户上门提出需求)的服务模式。

在大数据时代,我们必须要有这样的思维:传统数字图书馆的一切都要变,变得更加适应大数据,能充分运用大数据,把数字图书馆完全

融入大数据之中,促进数字图书馆在人们的学习、生活、工作和研究中发挥更大作用。

## 2 数字图书馆的大数据思维

数字图书馆的大数据思维,即指从大数据的角度考虑数字图书馆的各类问题,把数字图书馆完全融入大数据之中,增加数字图书馆数字产品,提升数字图书馆服务水平,借助大数据技术解决数字图书馆有关问题,把数字图书馆作为“互联网+”的重要分子。

### 2.1 数字图书馆资源建设

大数据时代,数字图书馆的资源建设思路应当转变。应当扩展数字图书馆的资源范围,把数字图书馆资源与整个互联网资源结合为一体,强调面向解决复杂问题的资源整合、资源加工的建设思路,树立大数据时代的资源建设观。

#### (1) 拓展数字图书馆资源范畴

数字资源是数字图书馆的立足之本,是提升服务质量和水平服务的保证。长期以来,图书馆都非常重视资源建设,但基本限制在文献型资源上,如图书、期刊、报纸、学位论文、会议论文、专利、标准、科技报告等。大数据时代,必须开拓思维,数字图书馆不仅仅是将上述文献数字化或增加一些数据库资源,还要将一些政府信息、社会关注信息、网络热点信息囊括其中,更需要将用户使用数字图书馆、搜索网络的行为信息作为数字图书馆的采集资源。另外,还应增加对再生资源的生产和汇集和存储,包括经过关联、挖掘分析后形成的各类综合数据,真正扩大数字图书馆的数据资源范畴。

#### (2) 增加数字图书馆资源整合的广度

大数据的特点是数据复杂多样,单纯文献信息已不能满足数字图书馆用户的需求,以文献服务为主的资源整合必然向以综合信息服务为主的资源整合发展。为了满足用户的各类需求,我们需要将文献信息、政府信息、社会信息、网络信息、用户信息进行整合,即将数字图书馆

资源与社会资源有机关联,尤其是将公众关心问题的社会资源整合到数字图书馆中。

另外,还需加强面向复杂问题求解的资源整合。大数据环境下,用户对数字图书馆的期盼,不仅仅限于文献的需求,更希望数字图书馆能够帮助他们解决实际问题,希望能够指导他们解决复杂问题。因此,那种简单的堆积和分类式的信息整合受到极大挑战,只有从决策角度出发,面向复杂问题的资源整合,才能真正满足用户需求。

#### (3) 加强数字图书馆资源组织加工深度

资源的组织加工是对资源的一种整序,其成果是一种资源的知识展现形式。在数字图书馆中,结构化信息(如书目信息)本身具有一定的知识表现形式,但这种知识表现基本上限制在资源库内部。在大数据环境下,我们必须拓展思路,努力将数字图书馆内外的资源建立语义关联,构成面向全社会资源的数字图书馆资源组织架构。

另外,在对图书馆资源的深度加工方面,应能够将信息经由知识解构的加工再深入至知识建构的加工。例如,从文献中分解出知识单元(知识解构),反过来将这些知识单元间或与文献间建立语义关联(知识建构),从而产生新的知识。在知识解构与建构上,除了强调利用数据挖掘技术、软件分析工具外,要特别提升图书情报领域所建主题词表、分类词表的应用能力和共享能力。

在数字图书馆资源组织加工方面,还有一种非文献资源必须得到重视,那就是用户行为信息资源。用户行为信息是数字图书馆知识服务与知识推荐的宝贵资源,需要深度挖掘、有机组织,做到将用户行为数据与文献资源、目标资源以及其他相关资源密切关联起来,使其成为知识服务的高效资源。

### 2.2 数字图书馆技术应用

大数据时代,数字图书馆技术已从处理局部数据转到处理更广域数据。从整个数字图书

馆的技术体系来看,包括数据采集、信息处理、组织架构、知识挖掘、分析预测、结果呈现、服务技术等。如何将大数据相关技术应用并融入数字图书馆领域,是数字图书馆领域必须要思考的问题。

#### (1) 语义技术

大数据环境尤其需要语义技术,如何使得大量复杂数据建立有机联系,需要靠语义技术来实现。数字图书馆要思考如何将词典(主题词表、分类表等)中的语义自动融入到数字文献相关信息中。当然,如此大量的数据依赖人工标注信息间的语义关系是不现实的,必须要借助有关词典,应用人工智能技术、本体技术、语义分析技术等,自动标注数据间的语义关系,使数据间充满语义联系,从而促进所收录资源的知识扩充和知识挖掘。

语义技术在图书情报领域的应用并不是新奇的事情,图书分类法、汉语主题词表等工具本身就是语义关系构筑起来的。但这些工具只帮助我们构筑了文献间的语义关系,很少涉及文献或信息内部知识点间语义关系的建立,然而在大数据环境下对这一点却提出了要求。知识点间的语义关系方便用户的知识获取,建立了通过某一线索借助语义关系获得有用知识的途径。

#### (2) 数据聚类技术

聚类是把相近的、有关联的信息或数据聚集在一起的过程。在繁杂、巨量的数据中,聚类是信息高效利用的有效手段之一。在数字图书馆中,更需要充分运用聚类技术,经过聚类的信息资源,将会在信息服务、信息分析和信息利用中发挥更大的作用。聚类不仅仅是对文献信息资源按照某一属性或特征聚集,还可以应用于用户需求,也可以运用于用户检索行为处理和分析。

大数据环境下,真正被利用的数据是很少一部分,聚类技术为这小部分数据的充分利用提供很好的途径。例如,对数据资源的聚类可将它们划分成一个个相互关联、主题相近的小数据集,这些小数据集非常适合面向问题

的需求,方便用户对信息的检索、选择和分析;对用户需求的聚类有助于我们对相近需求的对比分析,帮助用户优化需求,促进其充分表达需求;用户行为的聚类可以发现用户利用资源的行为规律,为用户个性化推荐服务提供数据支撑。

#### (3) 信息分析技术

大数据时代的数字图书馆应当能够充分利用信息分析技术。将大数据分析技术融入数字图书馆,并将数据分析作为数字图书馆的一项拓展工作,把分析结果作为向用户提供服务的高端产品。在这些分析技术中,如文献的信息汇编技术,可以为前沿研究领域的跟踪提供支持;联机分析技术(OLAP)可以在线分析用户利用数字图书馆的各类情况,为数字图书馆的管理运作提供数据支撑;还有许多定量定性分析工具与技术,如Hadoop、SPSS、CiteSpace等都应当在数字图书馆得到充分应用。

#### (4) 检索技术

未来数字图书馆的检索技术不能只限于本机构数字图书馆的检索,必须运用网格检索技术实现跨平台、跨资源的无缝检索。这需要构建各数字图书馆间的检索网络,每个数字图书馆能够实现馆内外资源关联检索,并提供网络资源搜索。学术搜索也应成为高校及科研院所数字图书馆提供的服务产品,以促进本单位学术研究。由于数字图书馆的利用与检索基本为自助方式,所以其检索界面应当通俗易懂,而且需要有更强大的功能,如语义检索、自动理解语言的对话式检索、跨语言检索等。

### 2.3 数字图书馆的产品与服务模式

大数据时代,文献信息已经不能完全满足用户需求,用户需要更加多样化的信息,不仅需要正式出版的文献信息,还需要视频、图像、非正式出版的灰色信息以及网络信息,更有甚者希望获取再生信息。

#### (1) 数字图书馆服务产品

学术资源建设商的产品类型越来越多,不

仅包括数字图书馆的几乎所有资源,而且还在不断扩充,对数字图书馆的影响越来越大。因此,未来数字图书馆的产品应该丰富多彩,才能更具竞争力。数字图书馆除了提供一次文献、有关网络信息的信息服务外,还应能够提供再生信息的服务,如知识库、方法库、推理库、战略库等智库产品。除此以外,科普教育文化类的视频信息也应成为数字图书馆的重要产品,如百科知识、科普讲座视频资料、历史文化视频讲座、大中小学精品课程视频等。

### (2) 数字图书馆分析产品

大数据时代不缺信息,缺的是解决复杂问题的知识,更缺少对众多信息处理分析、总结归纳后得到的分析信息,这类信息分析产品应作为数字图书馆未来信息服务的高端产品。因此,数字图书馆应注重用户行为分析成果,以帮助了解用户的需求,掌握用户关注主题和学术领域,并结合前沿领域的信息采集和科学分析,确定前沿领域和科学发展趋势。另外,通过用户行为分析也能够将用户经验进行总结推广,还可以发现用户的一些特异行为。对资源利用分析成果,可以了解资源的价值所在,可以分析资源的缺失情况,还可以得到有关资源利用规律等。

### (3) 数字图书馆服务模式

从传统图书馆到数字图书馆,其服务模式已由被动提供服务为主的模式逐渐转变为自助为主的服务模式。然而,在大数据环境下,这种自助式服务使许多用户虽处在信息的海洋中,却感到迷惘,常常不能得到(更不用说及时得到)自己最需要的信息。因此,数字图书馆必须改变过去被动式服务和等待式服务的模式,应采取主动推荐式的服务模式。如,根据用户行为有针对性地主动推荐,增加对热点信息或事件的重点推荐,强化定题信息的服务,等等。

另外,数字图书馆的数字参考咨询服务应得到更加快速的发展,未来的数字参考平台不仅仅是数字化,更要智能化。如,将自动应答、用户互答、专家解答相结合,相互间实现自动转

换的无缝衔接,使用户切身感受到参考咨询就是自己工作、学习和科研的良师益友。在信息类型的提供方面,数字图书馆应由文献提供为主的模式向多元信息提供的模式转换,如文化遗产、案例信息、综合分析信息、政府信息、声像视频信息,甚至城市社会信息等。

## 2.4 数字图书馆定位

大数据时代的数字图书馆应突破传统的图书馆思维,无论是数字图书馆架构,还是数字图书馆的服务理念以及工作重点,都必须有大数据思维。这就要求每一位数字图书馆馆员必须适应大数据时代,勇敢迎接挑战。

### (1) 跳出传统图书馆资源框架

大数据把我们带到了一个广阔的数据空间。数字图书馆面对的不仅仅是图书、期刊等文献信息,还面对着网络上种类繁多的信息。用户对数字图书馆的各类需求,促使对数据的采集、处理、组织以及服务都发生了很大的变化。我们必须构筑新的数字图书馆资源框架,要有全数据的理念,构建数字图书馆全数据框架。即数字图书馆不仅仅限于文献资源,还应把整个网络资源纳入数字图书馆框架体系,其框架内资源不仅仅限于文本信息,还包括图像视频信息。因此,信息的采集除了传统的订购方式以外,还要采取针对网络的信息自动获取手段;在信息的组织方面,应建立具有语义关系的、数据间联系更加紧密的信息组织框架。

### (2) 建立全方位服务理念

传统图书馆的服务基本是文献型服务,提供的服务形式多为等待型、被动型的服务。数字图书馆的服务资源已经有了拓展,各类数据库大大丰富了图书馆资源;服务形式也有了不少改观,由被动式的服务转向了自助式的服务。但从大数据时代这一角度出发,数字图书馆的服务理念还需进一步提升,应建立全方位服务理念。其一,面向全社会服务,即高校和科研院所建立的自有版权的数字图书馆内容也应面向全社会服务;其二,提供全资源服务,即提供各

种类型信息资源的服务,如文献信息、数据库信息、网络信息、视频信息、分析信息、政策法规、政府信息等;其三,服务方式和传播形式也要进一步拓展,即除了提供用户的自助服务和简答资讯以外,还需加强推送服务、个性化服务,并建立虚拟参考咨询平台,除了提供面向固定网络的服务,还需提供面向移动网络的服务,全面拓展数字图书馆服务。

### (3) 拓展工作重点

大数据时代,数字图书馆的工作重点要有所拓展,除了把握数字资源建设这一基础工作外,还要提升自己的服务能力。除了继续做好常规服务以外,还需定位一些高端服务,努力把数字图书馆建设成信息分析策源地;密切关注学科前沿领域,同时对国家重大科研计划领域或本单位重点科研领域进行跟踪,并提供科研进展报告,对政府行政、科学决策、科研攻关等提供数据支持。

### (4) 数字图书馆馆员定位

大数据时代对数字图书馆馆员也提出了更高的要求。每一位馆员应具有对数据的理解分

析能力,并立志成为数据分析师;要对大数据技术有一定了解,不一定要要求每一位馆员创造技术,但应当学会使用这些技术,尤其需要充分了解大数据运用于数字图书馆的关键技术;能够了解并熟练运用各种数据分析工具和软件;对网络资源能够全面把握,不仅仅是文献数据库,各类数据库都要知会,成为学术资源、网络资源的百事通。

## 3 结束语

大数据开启了数字图书馆的一次重大时代转型,将拓展数字图书馆的数据资源,提升数字图书馆的能力,增强数字图书馆的服务功能,丰富数字图书馆的服务产品,一个崭新的数字图书馆将会出现在我们的面前。在这样的新环境下,我们应该对数字图书馆有新的理解,在资源框架、技术应用、服务模式等多方面进行思考,用大数据的思维看待数字图书馆的变革,使未来数字图书馆在大数据环境下,在与学术资源建设商的竞争中,相互依存,共同发展。

## 参考文献

- [1] 中国知网资源总库[EB/OL].[2015-07-01]. <http://epub.cnki.net/kns/subPage/Total.aspx>. (China national knowledge infrastructure[EB/OL].[2015-07-01]. <http://epub.cnki.net/kns/subPage/Total.aspx>.)
- [2] 万方数据知识服务平台[EB/OL].[2015-07-01]. <http://www.wanfangdata.com.cn>. (Wanfang data knowledge service platform[EB/OL].[2015-07-01]. <http://www.wanfangdata.com.cn>.)
- [3] 超星数字图书馆介绍[EB/OL].[2015-07-01]. <http://baike.baidu.com/view/121194.htm>. (Superstar digital library[EB/OL].[2015-07-01]. <http://baike.baidu.com/view/121194.htm>.)
- [4] 百度介绍[EB/OL].[2015-07-01]. <http://baike.baidu.com/view/262.htm>. (Baidu[EB/OL].[2015-07-01]. <http://baike.baidu.com/view/262.htm>.)
- [5] 盘点中国2013行业数据量[EB/OL].[2015-07-01]. <http://developer.51cto.com/art/201402/430185.htm>. (Industry data quantity of China in 2013[EB/OL].[2015-07-01]. <http://developer.51cto.com/art/201402/430185.htm>.)
- [6] Huwe T K. Building digital libraries: big data and the library: a natural fit[J]. Computers in Libraries, 2014, 34 (2): 17-18.
- [7] Renaud J, Britton S, Wang D, et al. Mining library and university data to understand library use patterns[J]. The Electronic Library, 2015, 33(3): 355-372.
- [8] Borgman C L, Wallis J C, Mayernik M S, et al. Drowning in data: digital library architecture to support scientific



- use of embedded sensor networks[C]//JCDL '07: Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries, 2007.
- [9] Williams K, Wu J, Choudhury S R, et al. Scholarly big data information extraction and integration in the CiteSeerx digital library[C]//Data Engineering Workshops (ICDEW), 2014 IEEE 30th International Conference, 2014: 68-73.
- [10] 曾建勋. 数字图书馆的大数据[J]. 数字图书馆论坛, 2014(4): 卷首语. (Zeng Jianxun. Big data of digital library[J]. Digital Library Forum, 2004(4): preface.)
- [11] 陈传夫, 钱鸥, 代钰珠. 大数据时代的数字图书馆建设研究[J]. 图书情报工作, 2014(7): 40-45. (Chen Chuanfu, Qian Ou, Dai Yuzhu. Study on the construction of digital library in the age of big data[J]. Library and Information Service, 2014(7): 40-45.)
- [12] 刘炜, 夏翠娟, 张春景. 大数据与关联数据: 正在到来的数据技术革命[J]. 现代图书情报技术, 2013(4): 2-9. (Liu Wei, Xia Cuijuan, Zhang Chunjing. Big data and linked data: the emerging data technology for the future of librarianship[J]. New Technology of Library and Information Service, 2013(4): 2-9.)
- [13] 陈臣. 大数据时代基于个性化服务的数字图书馆数据搜索引擎设计[J]. 图书馆理论与实践, 2015(4): 91-93, 94. (Chen Chen. The search engine of digital library based on personalized service in big data era[J]. Library Theory and Practice, 2015(4): 91-93, 94.)
- [14] 陈茫, 周力青, 吕艳娥. 大数据时代下的图书馆移动服务创新研究[J]. 图书与情报, 2014(1): 117-121. (Chen Mang, Zhou Liqing, Lü Yan'e. Research about library mobile service innovation in big data era [J]. Library and Information, 2014(1): 117-121.)
- [15] 陈臣. 一种基于新型存储的数字图书馆分布式大数据存储架构[J]. 现代情报, 2015(1): 100-103. (Chen Chen. A digital library distributed big data storage architecture based on new type of storage[J]. Journal of Modern Information, 2015(1): 100-103.)
- [16] 李白杨, 张心源. 数字图书馆建设中大数据问题初探[J]. 情报科学, 2013(11): 26-29. (Li Baiyang, Zhang Xinyuan. Preliminary study of the big data mechanism in digital library construction[J]. Information Science, 2013(11): 26-29.)
- [17] 郭春霞. 大数据环境下高校图书馆非结构化数据融合分析[J]. 图书馆学研究, 2015(7): 30-34. (Guo Chunxia. The integration analysis of unstructured data of university library in big data environment[J]. Research on Library Science, 2015(7): 30-34.)
- [18] 王宇鸽, 邢军, 张红. 数字图书馆大数据技术的应用架构[J]. 图书馆学刊, 2014(7): 113-116. (Wang Yuge, Xing Jun, Zhang Hong. The application architecture of big data technology in digital library[J]. Journal of Library Science, 2014(7): 113-116.)

苏新宁 南京大学信息管理学院, 江苏省数据工程与知识服务重点实验室, 教授, 博士生导师, 教育部长江学者特聘教授。江苏 南京 210023。

(收稿日期: 2015-07-16; 修回日期: 2015-07-29)