

# 大数据在图书馆的应用研究与探讨

刘春颖

(长春市图书馆, 吉林 长春 130021)

**摘要:** 大数据作为一项新兴的信息技术, 对图书馆的数据存贮、数据分析计算能力是一次检验, 并会对传统图书馆业务产生不小的冲击。对于图书馆本身可以通过重视用户数据与信息、利用数据分析工具提高自身服务的创新应对考验与挑战。笔者对中美图书馆大数据应用现状做了对比, 分析并阐述了大数据给图书馆带来的机会与影响, 并对大数据在图书馆中的应用提出了建议。

**关键词:** 大数据; 图书馆; 服务

**中图分类号:** G250

**文献标识码:** A

**文章编号:** 1002-1248 (2014) 11-0090-03

## Big Data Applied Research and Discussion in the Library

LIU Chun-ying

(Changchun library, Changchun 130021, China)

**Abstract:** big data as a new information technology, it comes to the library of data storage, data analysis and calculation ability is an inspection, and will have a big impact on the traditional library business. For the library itself through attaches great importance to the user data and information, using data analysis tools to improve their service innovation service test and challenge. This paper made a comparison about the present situation of the library of China and the United States big data applications, analyzes and expounds the opportunity and influence of big data to the library, and the application of big data in the library are proposed.

**Keywords:** Big data; Library; Service

自从2012年起, 一个新的信息科技热词——大数据 (Big data) 的概念在全球被炒得比云计算、物联网还要火。种种迹象表明, 一个大数据时代正在到来。那么究竟什么是大数据呢? 一般来说, 大数据泛指巨量的数据集, 人们没法处理或者很难处理的数据量都可以叫大数据, 因为可从中挖掘出有价值的信息而受到重视。目前国际上大数据并没有标准定义, 其标准和产业格局尚未形成, 通常用“4个V”来描述大数据的特征<sup>[1]</sup>。

其实“Big data”在一些领域存在已有时日, 如: 物理学、生物学、环境生态学……只因为近年来互联网和信息行业的发展而引起人们关注。Internet的发展, 加快了信息化向社会日常生活的渗透。据全球最早提出“大数据”时代到来的知名咨询公司麦肯锡称: “数据, 已经渗透到当今每一个行业和业务职能领域, 成为重要的生产因素……很有可能, 不久以后, 毕业生找工作除了必备外语和电脑技能外, 还需具备第三样——数据素养”<sup>[2]</sup>。作为大数据时代具有3个特点:

(1) 在大数据时代, 主要应了解数据之间的相关性, “What is this”比“why”重要; (2) 在大数据时代, 研究与分析某个现象时, 使用的是非抽样数据而是全部数据; (3) 在大数据时代, 要适应数据的丰富与多样性。大数据时代的到来, 使图书馆由原来藏书楼、阅览室传统模式转型到只需一根网线、一个网卡即可轻松解决所有阅读需求的大数据数字图书馆时代, 快捷方便的电子书、立体生动的移动设备等使传统图书馆领域的旧有模式焕然一新。

## 1 大数据对图书馆产生影响的同时带来了发展机会

### 1.1 图书馆借助大数据建立和完善新的知识服务方式

在如今社会, 读者不断增加的信息需求促使图书馆的基础服务体系得到一定加强。图书馆运用大数据技术对用户的数据进行分析, 不仅可以了解用户的信息行为, 还可以深度挖掘用户潜在需求数据。通过分析读者构成、阅读习惯、潜在需求等借阅数据, 从而有针对性地开展对读者服务。例如, 一位读者近一段

收稿日期: 2014-03-27

作者简介: 刘春颖 (1973-), 女, 本科, 中级职称, 长春市图书馆。

时间经常借阅经济类图书,当他登陆数字化借阅端口,就会显示类似经济类图书供这位读者参考,每个人的推荐目录都是独一无二的。此项服务可以让图书馆更有目标性地增加主馆、各区图书馆和24小时图书馆的新增图书种类,打造实体图书和数字化图书为一体的新型阅读空间。

### 1.2 图书馆借助大数据建立知识服务导航机制

大多数读者都可能遇到过这种情况:走进图书馆借阅室,想要在分门别类的书架上找到自己需要的书有时并不是一件容易的事。遇到这种情况读者往往只能选择向管理人员寻求帮助,但多数情况下得到的帮助也不是感性的。这种情况下如果借助大数据建立知识服务导航机制就能很好的解决这个问题。例如利用移动应用程序将OPAC与馆内地图有机整合,就可以使读者方便地找到所需的图书与资料在书架上的位置。

### 1.3 大数据是图书馆拓展公共服务的技术支撑

数字图书馆(DL)、信息共享空间(IC)、IFRD、机构知识库(IR)、移动图书馆(ML)、云计算,提升了图书馆自身的服务水平和社会价值。其中人们较为熟知的“云计算”是大数据的基础平台与支撑技术。借助这些技术,图书馆可以从结构化和非结构化数据资源中抓取有用的知识和关联关系等,完善新的知识服务方式。大数据技术可以帮助图书馆获得准确、及时的用户数据,便于大数据的管理。建立优质的数据库系统并构建高质量的索引,了解图书馆最新知识服务的趋势并作出决策、调整服务方向,为读者提供高效率的数据检索功能。

### 1.4 大数据时代图书馆的数据存贮和计算能力面临挑战

大数据时代,数据类型不断增多,数据量急剧增大。用户的社会网络活动产生出大量的结构化和非结构化的数据信息,其数据的格式、类型、结构及存在形态等都变得更加多样,这就增加了大数据的复杂性。大数据对于图书馆的数据存储能力和范围、计算能力有着特别高的要求,必须要能对海量的数据加以分析,从中发掘出有价值的信息。图书馆虽自身拥有大量的数据,但其存储和运算能力受限,与大数据对存储能力的高要求还存在一定的差距。

### 1.5 大数据为图书馆建立新的评估标准

随着大数据时代的来临,人们对智能信息的需求更加迫切,数字图书馆就是在这样的历史条件下应运而生的。众所周知,数字图书馆具有信息资源数字化、信息传递网络化、信息利用共享化、信息提供知识化、信息实体虚拟化等特点。因而,评估图书馆标准,不再是衡量馆藏,而是其信息资源的获取、处理能力。

## 2 中美图书馆大数据应用

美国著名的《华尔街日报》把大数据时代、智能化生产、无线网络革命并称为未来繁荣的三大技术变革。

### 2.1 美国现状

大数据的重要性已经得到了世界各国的共识。据相关报道,美国已有大学专门开设了研究大数据技术的课程,培养下一代的“数据科学家”,一些美国公司也在向大学提供研究资助。2012年3月,美国奥巴马政府宣布推出“大数据的研究和发展计划”,5月,美国政府又发布了《大数据研究和计划》,用此项带有2亿多美元资金的倡议推动全民总动员来应对“大数据”时代的挑战,该计划也涵盖了图书馆部分。

### 2.2 国内现状

笔者在CNKI中国知网中以“图书馆”和“大数据”为关键词检索出有10多篇近两年发表的,大多是研究大数据对图书馆服务、发展等方面的论文。虽说,国内图书馆在大数据研究方面相比美国起步较晚,但国内IT行业对大数据的重要性有着清醒的认识。据统计,中国大数据市场2012年全年还不到5亿人民币,但有IT机构做了权威预测:未来几年国内大数据市场会突破百亿人民币。2012年,在东莞图书馆年会上,武汉大学的陈传夫教授展示了大数据时代数字图书馆体系结构图,并详细介绍了大数据对数字图书馆体系结构的改变。笔者工作的长春图书馆也于2012年5月31日开通了数字图书馆,读者只需使用阅读卡登录,就可以免费享受海量文化资源。2013年在深圳第十届“数字图书馆前沿问题高级研讨班”上,进行了主题为“大数据环境下的知识组织与服务——挑战与机遇”的热烈讨论。同年12月,武汉图书馆首次使用“大数据”分析全年图书借阅排行榜,发现读者阅读偏好从浅阅读向更高品质的充电式、启发式阅读和名著赏析变化……通过这些说明大数据已成为中国图书馆界的研究热点,可以简单了解大数据的应用状况。

## 3 图书馆应用“大数据”的策略建议

### 3.1 对大数据的存贮、分析、处理能力应提高

其实数据不一定非要用阿拉伯数字来记录,如:文字、音频、视频都是其常见格式。图书馆的大量数据应分为两部分存储,一部分考虑到光盘存储对数据存储本身而言具备离线存储功能,稳定,不易遭到破坏与攻击,以光盘的形式存储;另外考虑到硬盘存储速度更快、联网更方便,以硬盘形式存储。

以往,图书馆的数据处理牵涉的大部分是结构化的数据信息,可供分析的数据比例不大,要想从中挖

掘出十分有价值的信息难度较大。大数据时代的到来,其特征与优势促使读者的信息化需求发生变化,用户迫切需要图书馆提供智能化、个性化的服务。图书馆应及时改变数据处理范围和方式,利用现有条件挖掘潜力,对读者的借阅记录、信息行为、微博等各类数据进行深入分析,挖掘出有价值的信息。改善基础设施,提高服务方案,提高对大数据的分析和处理能力,提高服务效率。

### 3.2 图书馆面临硬件基础设施和技术人才的挑战

大数据时代,图书馆需要拥有经济、高效的存储和计算能力来存储和分析各类读者用户的信息,这需要建立在拥有较先进、完备的硬件基础设施和信息人才的基础之上。

### 3.3 图书馆应用大数据存在的隐私问题及解决办法

随着互联网的发展,读者在不同的地点会留下越来越多的数据痕迹。由于这些数据具有关联性、累积性,人们多关心其隐私可能被暴露。所以大数据时代图书馆应高度关注和重视大数据成本问题与隐私问题,进行数据分析和数据挖掘应该在不暴露用户个人隐私的前提下,坚持保护用户的个人隐私权。既要提倡数据共享,又要防数据被滥用。虽然全国人大已经通过加强网络信息保护的決定,使大数据的挖掘与利用有法可依。但还是希望早日出台“信息公开法”以适应大数据时代界定数据挖掘、利用的权限和范围;另一方面图书馆员自身应该树立良好的职业道德,在工作中高度关注和重视大数据的隐私问题,坚决维护用户的隐私权,这样才可以使读者的个人信息在合理、合法的范围内有效传播,以达到在充分发挥大数据优势的同时,又不侵犯用户隐私的共赢目的。

### 3.4 应用大数据图书馆应构建网络安全整体架构

Web 服务作为图书馆数字化的重要组成部分,它为最终用户提供了访问图书馆资源的重要接口。但是由于安全定位不同,使得传统的 IPS 或防火墙无法真正做到完全理解 HTTP,不能有效地保护 Web 应用业务安全。图书馆应采用专业级的 Web 防火墙对 Web 网站进行 Web 应用安全防护<sup>[3]</sup>,来阻断如 SQL 注入、远程文件包含、脚本……, schema 中毒、cookie 中毒、以及其他 Web 应用攻击的威胁,保护这个重要的入口。这样就可以保证 Web 应用程序安全性的同时还可防止敏感的数据库内容外泄,为图书馆 Web 应用提供了专业级的应用安全防护。

## 4 大数据时代图书馆服务与功能发生的变化

### 4.1 实现传统服务方式的改变

在大数据时代,数字图书馆、移动图书馆服务将

更加成熟。读者可以 24 小时内随时随地利用手机、平板电脑、掌上电脑等设备平台便捷检索、查询所需资料,既方便了读者又节省了大量时间,也为国内偏远地区不方便进馆阅读的问题找到一个解决办法。

### 4.2 可以实现针对不同读者进行个性化服务

大数据的应用就在于加强用户研究与交互数据的利用,并基于对用户数据的分析,提升个性化服务的水平<sup>[4]</sup>。各位网友在浏览网页时会经常遇到“猜您还喜欢……”,其实这即非结构化数据。是网站分析了用户的“浏览轨迹”,总结习惯共性而有针对性地提供推荐,这也给图书馆服务提供了一个新的思路:可以根据每一次信息服务的具体内容及用户长期的借阅记录,分析挖掘其数据模式及特征,提取有价值信息建立用户模型,实现对不同读者提供有针对性、个性化的服务。

### 4.3 分析大数据为科研机构提供研究新热点

在数据爆炸性增长,新类型数据不断涌现,结构更趋复杂的大数据时代。图书馆对拥有的大数据,可以借助相关软件进行分析,通过对大数据的分析为科研机构的决策服务,以此了解相关领域科研热点,这样既提升了图书馆在科研领域中的作用又对自身数据挖掘、分析能力起了促进作用。

## 5 结束语

近年来,信息技术的发展,数据量飞速增长,数据的存储与处理成本显著降低。数据挖掘已经渗透到了人们生活的各个方面,因此,必须要注意科学规划,从战略上重视大数据的开发利用。“热门话题”常需要“冷静思考”,大数据带来的不单是个大机遇,更是大挑战——大数据时代的信息安全需要重新定义。可以说,大数据时代对人类的数据驾驭能力提出了新的挑战,也为人们获得更为深刻、全面的洞察能力提供了前所未有的空间与潜力。在如今,大数据充斥在计算机、手机、监视器以及所有的电子设备上,连接了各类人群。不能否认传统图书馆是文明传播的主载体,而大数据将为图书馆的文明传播插上一对强有力的翅膀。

### 参考文献:

- [1] 张文彦.大数据时代的图书馆初探[J].图书与情报,2012,(6):15-20.
- [2] 韩翠峰.大数据带给图书馆的影响与挑战[J].图书与情报,2012,(5):37-40.
- [3] 杨海燕.大数据时代的图书馆服务浅析[J].图书与情报,2012,(4):120-122.
- [4] 王捷.大数据时代下图书馆开展信息服务的对策[J].现代情报,2013,(3):81-83.