

【编者按】本栏目由本刊和e线图情 (www.chinalibs.net、www.chinalibs.cn) 合办,对图书情报领域的知名专家进行特约访谈。受访专家就特定问题的真知灼见将陆续见刊,以飨读者。本期受访专家北京大学信息管理系主任、博士生导师李广建教授。研究方向为信息资源管理、网络信息管理技术与信息系统。

访北京大学信息管理系主任李广建教授

刘锦山

(北京雷速科技有限公司,北京 100085)

二十多年前,数字图书馆的兴起为图书馆学、情报学创造了一次良好的发展机遇;现在,大数据的出现似乎又一次为图书馆学、情报学的发展提供了新的机会。我们应该如何看待新技术发展对于图书馆学、情报学的影响?如何把握这种机遇以推进学术与实践的发展?带着这样的问题,e线图情采访了北京大学信息管理系主任李广建教授。

刘锦山:李老师,很高兴您能接受我们的采访。您多年来一直从事情报学与图书馆学的研究与教学工作,取得了丰厚的成果。请您首先向读者朋友谈谈您的学术、工作经历。

李广建:谢谢e线图情和刘总对我的关注。我的工作经历和学术经历比较简单。我的本科和硕士研究生都是在北京大学信息管理系念的,当时还叫北京大学图书馆学系。1980年入校,那时只有图书馆学专业,我是理科班。入学之后,系里将同学分成若干组,先到相关各系学习专业基础课和部分专业课,然后再回到图书馆学系,学习图书馆学的相关课程。我当时分在计算机组,在计算机系读了两年后,又继续回到本系学习图书馆学课程。1984年本科毕业,同年考取本系研究生,1986年年底硕士研究生毕业之后,被分配到北京师范大学工作。此后一直在北京师范大学教书,1999年至2002年在中国科学院文献情报中心攻读博士学位。2007年年底回到北京大学信息管理系任教。

刘锦山:我觉得北京大学图书馆学系当时将新入学的同学分成不同的小组很好,有利于人才培养。

李广建:是的,我的印象里,1979级和1980级是这么做的。至少对我来说,从这种教育模式中受益匪浅。在计算机系的学习确实开阔了自己眼界,计算机方面的知识对我以后的工作和实践都有很大帮助。

刘锦山:您的研究生导师是哪位老师?

李广建:我的硕士导师是张树华老师。我的博士生导师是徐引麓老师。北京大学和中国科学院两个地方的读书经历对我的帮助和启发很大。北京大学图书馆学系在学术方面为我提供了非常好的知识积累和方法论训练,而在攻读博士学位期间,中国科学院文献情报中心进入中科院知识创新工程,建设数字图书馆,我也因此在徐老师的指导下接触、参与了文献情报中心的一部分工作,这两段经历对我影响很大。特别重要的是,我的两位导师不仅在学术上给了指导,她们诚挚谦虚的品格和宽厚善良的处世方式,至今都是我的行为楷模。

刘锦山:李老师,您的硕士和博士论文做的是什么呢?

李广建:我的硕士论文题目是《论图书馆社会学》。1986年左右,图书馆学基础理论发展比较快,学

界特别关注图书馆与社会的联系,强调图书馆应该起到的社会作用。张老师的主要研究方向是基础理论和读者工作这两块,我在她的指导下选了这个题目,论文主要包括两个方面的内容:一是用社会学的方法去研究图书馆,二是强调社会和图书馆的联系。

博士论文的题目是《个性化网络信息检索系统的研究与实现》,主要是通过捕捉网络信息检索系统中用户的检索行为,进行用户建模,判断用户的学术兴趣,主动为用户推荐他们可能感兴趣的東西。

刘锦山:李老师,我们注意到,您先后在北京师范大学、中国科学院文献情报中心、北京大学等单位工作,既从事理论研究与教学工作,还参与了一些实践项目的工作,这样的学术与工作经历是十分可贵的。这样,无论理论研究还是实际工作,都可以转换视角,从不同视域来看待问题。您是如何看待您过去的学术、工作经历对您的影响的?

李广建:我在北京师范大学工作了十几年,一直从事教学科研工作。在中国科学院文献中心学习和兼职工作也有将近十年的时间,中国科学院文献情报中心从新馆搬迁,到数字图书馆建设,我都经历过并参与了其中一部分工作,收获确实非常大。同时我还在中科院文献情报中心兼任过一段博士生导师,参加博士和硕士的教学工作。这两种类型的单位,教学和科研有共性的方面,但各自的特色也非常明显。

在共性方面,这两种类型的单位都非常讲究教书育人,对教学有严格的要求,作为教师,要想搞好教学,自己必须要好学。给人一碗水自己就要有一桶水,教师也时刻要逼着自己去学习。

另一方面,在高校里,教学和科研都特别强调知识的系统性。教学上,比较强调学科的知识体系,教学过程通常要讲求基本概念、基本原理和方法体系。在科研方面,高校教师有比较大的自由度,一般有自己的稳定的科研方向,追求对学科知识的贡献。在中国科学院文献情报中心这样的实践单位,当然也有自由的科学研究,但更多的还是面向问题、解决实际问题的研究。来了一个问题,这个问题怎么解决,要结合问题搞研究,研究的目的是结果一定要落实到问题的解决上面。同样,在教学方面,无论是课程的设置,还是教学内容,也会反映出这种面向问题的特点,因此,毕业的硕博生解决问题的能力都比较强。

当然,以上仅仅是我个人的体会,不一定全面和准确。

我非常感谢这几个单位,北京大学培养了我,北京师范大学和中国科学院锻炼了我,在这几个单位的不同经历,对我的发展很有帮助,给了我书本上的理论与图书情报工作的实践两者结合起来的机会。

刘锦山:李老师,数字图书馆一直是您关注与研究的领域,请您谈谈您如何看待过去20年来数字图书馆的发展状况及其未来发展趋势和影响?

李广建:数字图书馆在我们国家也有近20年的发展历程了,取得了很大的成就。成就至少体现在两个方面:一方面由于数字图书馆研究的驱动带来了图书馆的理论和实践的变化,进而促进了图书馆学、情报学等学科的发展;另一方面数字图书馆本身的发展,促进了图书馆事业的发展,同时又培养了用户,激发了用户的需求。在我看来,用户的需求是需要激发的,用户需求有现实需求和潜在需求,数字图书馆的发展把用户很多潜在需求都激发出来,例如对电子资源的使用习惯的培养。我举个例子,在早期建设数字图书馆的时候,大家还是比较习惯去看纸本的东西,对电子资源还不太习惯,随着数字图书馆一步一步地向前发展,用户觉得数字图书馆的系统与服务方式越来越便捷,用户的使用习惯也不断在转换,反过来又对数字图书馆提出了更高的要求,这是数字图书馆很大的贡献和成就。过去,很多事可能想到了也做到了,但是有些事是想到了却做不到,因为没有技术手段或者做起来很复杂,就不做了。数字图书馆的发展使得以前这些做不到的事情现在可以做了,甚至有可能由于数字图书馆的发展,数字技术与用户需求的相互促进,很多过去没有想到的事,现在我们不仅能够想到,而且能够做到。再进一步说,数字图书馆目前正扮演着用户需求的主动引领者的角色,而不再是用户需求的被动响应者。这是数字图书馆的最大发展成就。

就未来趋势而言,数字图书馆的发展有这样几个问题值得大家关注:其一,数字图书馆的互联互通以及数字图书馆与互联网的融合。美国ALA曾界定了数字图书馆的若干特征,其中之一就是:数字图书馆不是一个实体。当前,在实践上,很多号称“数字图书馆”的系统,都还是基于实体来建设的,一个单位自己建一个数字图书馆,购买或自建一些资源,用户只能在这个系统中打

转转,系统和系统之间并没有真正的互联互通,从某种意义上来说,这些数字图书馆都是互联网上的信息孤岛,就更不用说真正地融入到互联网中去了。这里所说的融入到互联网,是指数字图书馆与互联网业界的合作。这种合作不是指传统的合作方式,不是指图书馆将自己的数据公布到互联网上去,如将馆藏数据公开到“谷歌学术”上去,而是指要充分利用互联网上的数据资源为我服务,同时将自己的资源开放给互联网上的其他系统。例如,利用用户在搜索引擎中查询学术信息的行为,预测用户的阅读兴趣,提升数字图书馆的服务主动性,远比利用一个数字图书馆的用户日志数据来得更为全面和准确。因此,我觉得互联互通、与互联网融合这可能是将来发展的一个趋势,最近几年开始的互联网数据开放,将为数字图书馆与互联网的融合提供发展机遇。例如,百度已经正式发布大数据引擎,将包括开放云、数据工厂、百度大脑三大组件在内的核心大数据能力开放,通过大数据引擎向外界提供大数据存储、分析及挖掘的技术能力,数字图书馆完全有可能在这种环境下有更大的发展。

其二,服务公平化是将来数字图书馆发展的大趋势。现在一说到数字图书馆,通常就想到“高”、“大”、“上”、面向科研这些东西,但是数字图书馆的初衷至少不是为了所谓的“高”、“大”、“上”的,因此服务的公平化是数字图书馆发展的目标。过去对数字图书馆的研究,比如研究用户行为方式,比较多地针对科研人员等高端用户。实际上,大部分的阅读,特别是数字阅读、电子阅读,可能更多的是普通老百姓,但目前我们对普通人员的电子阅读的研究并不多,以后可能需要更多地关注普通人的电子阅读习惯。数字资源的使用者可能不仅仅局限于高端用户上,普通用户也会越来越多。

其三,移动是数字图书馆的重要发展趋势。服务均等化与移动也有关联。这么多年来,数字图书馆的概念在不断发展,相关的新名词不断出现,例如“遍在图书馆”。“遍在图书馆”其实是数字图书馆的另一种称谓,但侧重点与“数字图书馆”有所区别,它更强调数字图书馆适应于不同的人群或者不同环境下的人群,比如不能读的可以听,不能读也不能听的可以摸。一个人,半夜想起来一件事,想上“遍在图书馆”去查查,在这样的情况下,可能光线不足,他可不可以不用开灯或者不用起床走到计算机的旁边就通过摸和听的方式来使用数字资源呢?这就是遍在图书馆要考虑的事情。在任何环境下都可以对资源进行访问,这是一个大的发展趋

势。移动可能会带来新的理念和方法,与传统、固定的网络不太一样,因移动而遍在。同时,移动还有可能带来一些新的数据采集方式和服务方式,例如,通过移动终端可以测量人的心跳情况,以后用户在用移动终端阅读的时候,移动终端会自动测出人们的心跳,根据心跳频率来判断用户的兴奋程度,从而知道用户对看到的内容是否很关注,用网上流行语来说,那个时候,点赞是一件很out的事情,因为,你心里想的是啥,数字图书馆早就知道了。

第四,跨界也是数字图书馆未来的重要发展趋势。过去的数字图书馆资源更多地还是限于传统文字型资源。近几年,不同媒体与不同服务的跨界成为一种趋势。以前的数字图书馆只提供一种资源的表现形式,即数字化的图书或文章。现在很多数据库都会把核心作者、相关文献、引文文献等列出来,服务都延伸了很多,甚至可以提供篇章、科学数据、实验数据之间的联系。在以前纸质论文中,有很多细节如实验过程等是没有办法详细地展现出来,现在,可以把科学数据、研究过程和科学文献打包,成为数字对象,读者除了能看到研究结果,还可以去看研究过程甚至是数据的细节。数字图书馆已经不是提供简单的文字阅读功能了。

刘锦山:李老师,您谈到数字图书馆不是实体这点很对。一般情况下,人们认识一个东西,总感觉实体的东西比较踏实、具体,很好去认识。但是如果总局限于这样一个阶段,总是把认识对象局限于实体,就会限制认识的进一步发展。科研的发展过程也是逐步从实体到普遍的概念。数字图书馆不是一个实体,当然数字图书馆包括实体的东西,但是我们不能把它仅仅当作实体来看,只有把这个概念理解了,才能更好地理解均等化的发展趋势,二者是密切相关的,只有普遍化了以后才能做到服务的公平和均等化。

刘锦山:李老师,最近几年,大数据引起了各界的重视。大数据应该是两三年以前美国先提出来的。同时,学科交叉也是现在学术科研发展的一个非常重要的方向。您对于图书馆学、大数据、情报学这三者均有涉猎和研究,请您从学科交叉的视角谈谈这三个领域之间的交集会给我们有些什么样的启示呢?

李广建:大数据出来以后我也一直在学习。如果说数字图书馆当年给图书馆学、情报学提供了一个机会,

那么大数据的出现又一次为图书馆学、情报学提供了一个机会。为什么这么说呢？

从现在来看，到底什么是大数据，尽管现在还有不同的定义，但认真细致地总结一下，关于大数据的概念有两种比较典型的看法，或者说是隐含着两种不同的看法。一种是强调数据及其处理技术，当然这是很重要的，早期的大数据概念比较强调这个特征。另一种看法是最近几年开始更多强调的是大数据、大洞察，简单来说就是大数据分析，现在大数据讲的是怎么从大数据的分析中得到有益的结果。

为什么说这又是提供了一次机会呢？以情报学为例，情报学中非常重要的领域就是信息分析和情报分析，文献计量、科技情报分析、竞争情报等都是我们这个领域中耳熟能详的名词术语，它们的共同特点是通过与信息或数据进行分析、从中提炼出知识并加以诠释，再映射到自身领域来解决或描述各种问题、现象或者提供决策支持。从这个意义来说，情报分析和大数据分析并没有本质的区别，大数据是不是可以看成是情报学的“回归”呢？另一方面，大数据的理念和方法在某种程度上又促进、丰富了情报学和情报分析的研究，过去，我们更多的是处理单一类型的数据，大数据的概念来了以后开阔了我们思路，例如，在情报分析领域，有关多源信息融合和多源方法融合方面的探讨越来越多。

还有一个机遇，这么多年来，大数据分析的理论方法到底是什么、方法体系到底是什么，在大数据研究和实践领域都没有公认或者成形的答案，也就是说，这个问题并没有彻底的解决。情报分析已经做了这么多年，是不是应该对此做点贡献？我的意思不是让情报分析去代替大数据分析，但是，情报分析是不是应该为大数据分析的方法论做点贡献？BDA (Big Data Analysis) 中的A不也是分析么？事实上，这本身对情报分析理论与方法的建设也有好处。情报学的学科发展肯定不能够固步自封，要吸收别人的养分而变得强大，同时结合自己优势为大数据的发展做出贡献。大数据不仅是一个概念，更多的是一个理念，实际上到底什么是大数据，多大才是大数据，并不是十分重要，我觉得理念、思想和方法论更有意义。

第三，大数据为情报学提供了实践机遇。刚才我们说在理论方面，情报学应该有贡献；同时，情报学应该抓住实践领域的机遇，大数据的发展为情报学的学科体系建设、人才培养提供了很好的机遇，怎样抓住机遇，应对需求，反过来促进学科的发展。所以我们真的

要去抓这个机遇。前一段时间我的一位同事讲，目前大家公认的几个世界上最牛的数据科学家都在以色列，而这几个数据科学家原来的背景都是搞情报分析的。别说是搞情报分析的，就是主流的数据科学家中有一两个出身于情报分析领域，就足以说明情报分析是有点用处的。我觉得这种影响真的是实实在在的，所以希望我们也可以抓住这个机遇，我只是从情报分析的角度来说。情报学分析包括信息采集、存储、分析、提供服务等环节，这种思想方法贯穿在整个大数据分析的流程里。

还有一个例子能看出来图书馆学、情报学与大数据之间的关系。我们系的毕业生有相当一批人在做与互联网和大数据处理有关的工作。如果有1个人、2个人做这方面的工作，可以说是个人的特质和兴趣所致，但如果做这方面工作的人比较多，是不是说明两件事的关联性比较大？我们希望在课程体系建设、人才培养、科学研究方面围绕大数据做些工作。

刘锦山：李老师，我们注意到，信息资源整合是信息资源深度开发利用的一种重要方式，但传统的整合方式多是基于资源属性的一种整合方式，其关注点并不在用户。针对这种情况，您提出了基于生活事件的信息资源整合方式，请您谈谈你这一研究成果及其意义。

李广建：首先，要说明的是，基于生活事件的信息资源整合方式并不是我提出来的，如果我没记错的话，最早是欧洲的学者提出的，我只是写文章介绍或引用罢了。在以前搞数字图书馆研究和参与数字图书馆建设的过程中，我发现，特别是早期的数字图书馆系统，基本上是建设单位的成果展，开发很多个系统最后都发布到网页上，图书馆买一个资源就放到网页上，读者一开始觉得挺新鲜，但是后来就烦了，因为，网页上的系统越来越多，每一个系统的规矩又都不一样，进去之后就要学习每一个系统的使用方法；面对众多的系统和资源库的入口，用户不知道所需要的东西在哪里，反而在丰富的系统和资源面前感到迷茫。我们一直在强调数字图书馆的建设应该以用户为中心，但实际上，许多数字图书馆建设还是以自我为中心，也就是以建设者为中心的。用户要想使用数字图书，就必须知道什么叫题名、文摘、关键词，什么叫组合检索、复杂检索等等，夸张一点说，就是强迫用户去学图书情报学，否则就不能很好地使用数字图书馆。

谷歌、百度为什么老百姓那么愿意用，就是因为界

面很简单,用户可以什么都不知道,想找的东西一输入检索词就出来了。再看看我们现在的检索系统,界面比较复杂,简单检索、复杂检索、全文检索等等一应俱全,对普通用户来说,增加了他们使用数字图书馆的成本。

基于生活事件整合信息资源的做法最早起源于电子政务领域。不管大政府还是小政府,都有很多不同的部门,各个部门有自己的职责范围,各自职能的范围都是各管一段,国外也是不可避免。原来我们总说国外办事效率高,那只是后台整合得好。对老百姓来讲,尤其是国外的老百姓,政府的概念很淡薄,不管政府设立什么部门,我要干的事就是我想干的事,你只要来告诉我该怎么做就成了,不用去学习民政部门干什么、人力资源部门管什么等。比如,我现在就想去就业,想要找工作,找工作可能要涉及一系列事情,包括提交受教育情况的证明、居住情况证明等等,教育、居住可能分属不同的政府部门管理,在非电子环境中,可能就需要求职者分别跑几个部门才能办好这些事情,而在网络环境下,用生活事件来表示公民(或企业)在生活(生产)中遇到的需要政府服务的生活场景,将用户需求映射为服务流程,用户无需知道他需要的服务来自哪个政府部门,而只需确定自己要干的事情,就可以被引导至相应的服务。换句话说,老百姓不必知道政府有哪些部门,就可以一站式地解决自己的问题。事实上,这种思想在数字图书馆、信息服务这些领域里面其实都有,有时是有意识地在做,有时就是无意识地做,在电子政务领域,这方面的研究和实践都做得比较突出,例如,欧盟前一段时间就资助了好几个这类的电子政务项目。

刘锦山: 李老师,北大正在改革博士招生方式,贵系将从2016年起将博士招生的考试制改为申请制,请您谈谈这方面的情况。

李广建: 这方面的变化是比较大的。国家已经批复了北大综合改革方案,改革有很多方面,关于博士招生这块,将由以前的考试制改成申请制。原来学校有一些试点的院系,我们是2014年把方案上报给了学校,将于2016年实施。

考试制有好的地方,当然也有问题,最大的问题就是一锤子定音。从总体上是公平的,可以选拔出人才,但是从个体上来看,并不一定能保证选拔出来的人就是最优秀的。申请制能够在对学生做全面了解的基础上进行选拔,可能会更加符合老师和学校对学生培养的要求。申请制的具体操作方式是这样的,首先申请者要有一定的基本条件,这些基本条件包括学习条件、科研条件、经历条件等,符合条件就可以申请。系里成立的委员会对申请者进行初筛、面试、能力测试这几个阶段来确定人选。

有人可能担心就是申请制会不会出现“走关系”的情况,产生招生腐败,这一点我认为不必担心。首先,我们在制定招生细则时,已经充分考虑到招生委员会和招生导师之间的权力分散和相互制衡,确保选拔出符合条件、科研能力强的申请者;其次,在今后的四年培养中,还有一系列中期考核、开题、匿名评审的环节保证培养质量,有多个环节过滤出不合格或不适应的学生,同时,学校对于有不能按期毕业学生的导师也有相应的奖惩措施,这些措施保证了导师在挑选学生时,会慎之又慎。

作者简介

刘锦山,男,1969年生,博士,北京雷速科技有限公司董事长,E-mail: ljs@leisun.net。

(收稿日期: 2014-12-05)