

中文电子书资源概述

王小宽

社科院 计算机网络中心

【摘要】电子书定义大致有三种：广义电子书—所有数字化出版物,有别于纸载体的传统出版物¹；狭义电子书—既有纸书的电子版；专门阅读电子书的掌上阅读器。本文涉及的电子书专指纸书的电子版。

千百年来书籍一直是人类社会文明信息传承的主要载体。数字革命正在迅速而深刻地改变着这种状况。「电子书」应运而生。近十余年来大陆电子书市场发展很快,经历了从纸质书籍简单复制和仅提供简单阅读查询逐渐发展到对海量电子书数据库整体信息进行整合并实现多重检索和深度信息挖掘。

大陆各类电子书情况综述,并从收藏规模、特点、侧重、收费标准、技术、格式、用户、数据挖掘、辅助服务等方面,对大陆具代表性的电子书提供商北京超星、方正阿帕比、书生之家进行比较评述。海外大型电子书项目简述,如GOOGLE、微软、联合国教科文组织牵头的全球数字图书馆、欧洲数字图书馆等。

当前电子书所面临的知识产权问题及解决之道。

电子书的出现使当代人类社会的信息环境发生了革命性的变化,目前还只是处在这场大变革的初级阶段。在可以预期的未来有可能出现被称为人类最后一本书—「电子书」,那时人们将真正做到所需要的信息知识「一点即得」—这是IT界、信息界、知识界的共同努力目标。

【关键词】中文电子书、数字图书馆、数字化

¹ 百度百科 http://baike.baidu.com/view/17637.htm?fr=ala0_1_1

千百年来书籍一直是人类社会信息知识传承的主要载体，在现代媒体出现之前，「书籍」就意味着历史、权力、知识、地位和财富，「书中自有黄金屋，书中自有颜如玉」。从某种意义上讲，进入文明社会，特别是近现代教育制度的形成，人的社会化过程也主要是通过「念书」来完成的。读书、教书、著书、购书、藏书、抄书、查书几乎成了学术活动的代名词。

古人将「读书万卷」作为知识渊博的一个衡量尺度。假定古代写在纸帛上的「一卷书」大致计为 5,000 字(17 字*311 竖行，也就是一部老子的《道德经》)，万卷书所含的信息量相当于 5,000 万字。古文言简意赅，折合成现代汉语大约为 1 比 5，合 2.5 亿汉字，以现代书籍每部平均 25 万字计，合 1,000 种著作。如此算来古人的「读万卷书」约相当于现代人「读千本书」，实属难得。毛泽东一生博览群书，有据可查超过 1,200 种²，实际当远超此数。在海峡两岸都享有盛誉的李敖，据说每天坚持 14 个小时的阅读，而且坚持了几十年，读书最多，号称中国第一³。至于到底读过多少，尚无统计。姑且按两天读一本计算，60 年大致有一万种，已是常人难以企及的境界了。

那么，古往今来的中文书籍总量到底有多少？

先说 1912 年以前的古籍⁴。有 10 万、15 万、20 万等诸多说法。李敖先生说有 25 万 3 千种，现存 10 万余种⁵。最新的说法是 20 余万种⁶。

1912 至 1949 年估计在 15 万种左右⁷。

1949 至 1978 年大陆出版新书 24 余万种；1979 至 1999 年出版新书 84 余万种；2000 至 2008 年出版新书 93 余万种；累计出版新书逾 201 万种⁸。

² 据中央文献研究室 2009 年出版的《毛泽东读书集成》，共 256 卷，达 3 亿字，收录了毛泽东阅读和欣赏的图书、文章、戏剧、书法 1,195 种，参照通行的图书分类法，划分为马克思主义、哲学、自然科学、社会政治、经济、军事、历史、文学、书法、报纸杂志、丛书工具书共 11 编。而毛泽东保存下来的藏书，有 1 万余种，近 10 万册。参见 http://www.leadermao.com/E_ReadNews.asp?NewsId=152。

³ 参见李敖《要把金针度与人——从《中国名著精华全集》谈怎样读书》。

⁴ 凡产生于 1912 年以前，并以稿本、抄本、印本、拓本等形式行世者，诸如简帛典籍、敦煌遗书、宋辽夏金元明清时期版印抄写的古籍、古旧地图、碑帖拓片、少数民族文字古籍，以及西学传入后产生的新学书籍等。国家古籍保护中心，《中华古籍总目编目规则》，2009 年 10 月。

⁵ 参见李敖《〈中国名著精华全集〉序》

⁶ 历经十余年编辑整理、即将出版的《全国古籍总目》，含现存古籍二十余万种（包括部分海外藏的中国典籍）。<http://www.guji.cn/opennews1.php?id=193>。

⁷ 据北京图书馆编《民国时期总书目》（书目文献出版社，1994 年）不完全统计，为十二万四千种。

⁸ 该数据不包括中小学课本和图片，参见 <http://www.chuban.cc/cbtj/dlcbtj/>；

台湾、香港澳门以及新加坡、日本、南北韩、越南，加上世界各地自古以来出版发行的中文书籍未见准确统计，应该也是海量数字。

据估计全世界有史以来共出版过书籍超过 4,000 万种⁹。

我们面临着这样一种困境：一方面累积的、以纸书为载体的信息总量已远远超出人们以传统方式处理信息的极限—或者受各种条件的限制无法接触到相关的书籍；或者面对浩如烟海的资料信息，叹人生苦短，不能物尽其用。而与此同时，「信息爆炸」依然以几何量级的方式在发展。

幸运的是，信息化正在迅速彻底改变着这一情形。以电脑网络为特征的数字革命正在迅速深刻地改变着既有的社会信息环境。以收藏「纸书」为主的图书馆在向虚拟的数字图书馆迈进，人们的阅读习惯在改变，人们对信息获取的要求在提高，电子书应运而生。

自上世纪九十年代中期开始，大陆众多的政府部门、图书馆、专业 IT 商家纷纷投入巨额人力物力构建以电子书为主要内容的数字图书馆。十余年走过来，超星数字图书馆、方正阿帕比数字图书馆和书生之家数字图书馆成为业中佼佼者。

北京超星公司¹⁰成立于 1993 年，主要从事图书的数字化加工工作，1997 年首家网上数字图书馆「瑞得—超星网上数字图书馆」。1998 年超星公司以技术优势，入选国家图书馆（原北京图书馆）数字图书馆建设项目，为其数字化加工了 17 万种中文电子图书。2000 年与广东省立中山图书馆合作，创办「超星数字图书馆」品牌。

超星公司目前已数字化完成 200 万种中文图书，其中已解决信息网络传播权的扫描类中文图书达 100 多万种，非扫描类中文图书达 30 多万种，2000 年以后出版的新书也已达 25 万多种，是目前全球最大的中文电子书资源提供商。

2008 年正式推出「读秀学术搜索」¹¹，据称是全球最大的图书、期刊搜索及

<http://www.gapp.gov.cn/cms/html/21/464/List-1.html>。

⁹ <http://baike.baidu.com/view/3223054.htm>

¹⁰ <http://www.ssreader.com/>

¹¹ <http://www.duxiu.com/>

文献传递系统。它集图书、期刊搜索、报纸、论文、视频、图书试读、期刊试读、文献传递、参考咨询等多种功能为一体，以海量的数据资源库为基础，为用户提供切入目录和全文知识点的深度检索，以及部分图书的全文试读。读者通过可直接进入各种检索的结果或某个章节进行图书、期刊的阅读，也可通过文献传递来阅读，帮助读者获取他们想要的文献资源，是一个真正意义上的知识搜索及文献传递服务平台。

北京方正阿帕比技术有限公司¹²（以下简称「方正阿帕比」Apabi）成立于2006年。Apabi 分别代表着 Author（作者）、Publisher（出版者）、Artery（流通渠道）、Buyer（读者）以及 Internet（网络）。作者、出版社、发行商和读者是传统出版产业链的有机组成部分。方正阿帕比以互联网为纽带，将传统出版的供应链有机地连结起来，实现完全数字化的出版。

中国 90%以上的出版社在应用方正阿帕比技术及平台出版发行图书，累计近 60 万种。中国 90%的报业集团、800 多种报刊正在采用方正数字报刊系统同步出版数字报纸。

方正阿帕比电子书及数字图书馆系统—「中华数字书苑」收录大陆出版的 20 万种、内容涉及政治、经济、历史、文化等各个领域的电子图书，受到政府部门的高度认可，多次作为馈赠国礼。方正阿帕比还推出「孔子学院数字图书馆」精选了数十家国内知名出版社合作的数万种中文电子图书，内容涵盖哲学、宗教、社会科学、政治、法律、军事、经济、文化教育、语言文学、艺术和历史地理在内的各个学科领域。

北京书生公司¹³创立于 1996 年，一直致力于以数字技术取代传统纸张应用，2000 年正式推出书生之家数字图书馆。它集成了图书、期刊、报纸、论文、CD 等各种载体的资源，下设中华图书网、中华期刊网、中华报纸网、中华资讯网和中华 CD 网等子网。资源内容分为书（篇）目、提要、全文三个层次，并提供全文、标题、主题词等十种检索功能。书生电子图书大部分为文本格式的全息电子图书。书生之家数字图书馆收录入网出版社 500 多家、期刊 7,000 多家、报纸 1,000

¹² <http://www.apabi.cn>

¹³ <http://www.sursen.com/>

多家，现有 2000 年至今的图书 30 万余种。

近年来，海外一些机构也相继推出数字图书馆项目，影响较大的有：

全球数字图书馆¹⁴，在美国卡内基梅隆大学教授 Raj Reddy 和浙江大学原校长潘云鹤等人的倡导下，2000 年联手筹建高等学校中英文图书数字化国际合作计划 (China-America Digital Academic Library, CADAL)，目标是完成一百万册中英文图书的数字化。其中中国（大陆）部分由浙大和中科院研究生院牵头，北京大学、清华大学等国内十六所重点高校共同完成。

2002 年 CADAL 演变成「中国大学数字图书馆国际合作计划」(China Academic Digital Associative Library)¹⁵。CADAL 与「中国高等教育文献保障系统 (CALIS)」一起，共同构成中国高等教育数字化图书馆的框架。2005 年底 CADAL 正式开通，定名为「全球数字图书馆」(Universal Digital Library, UDL)。

CADAL 一期收录的中文图书包括珍贵古籍、民国时期书刊、博士硕士学位论文等特色文献资源。英文图书则包括美国大学图书馆核心馆藏、技术报告等图书资料。截至 2007 年底「全球数字图书馆」馆藏已突破 150 万册。其中，中国（大陆）部分完成了 107 万册，占「全球数字图书馆」的百分之七十以上。

2009 年 8 月 CADAL 项目二期正式立项，由政府投入 1.5 亿建设资金，拟再增加 150 万册（件）数字资源，并建立分布式数据中心和服务体系，实现数据安全和全球服务，计划在三年内完成。印度、埃及、瑞典、德国等国已陆续加入「全球数字图书馆」项目，同时中国大陆参加 CADAL 项目的高校从原来的十六家扩增到近四十家。CADAL 还与多所世界名校签订了校际数字图书共享协议，力争早日实现馆藏一千万册的目标。

世界数字图书馆 (WDL)¹⁶以「促进国际和文化间的相互理解、增加互联网上文化内容的数量和种类、为教育工作者学者和普通观众提供资源、加强伙伴机构的能力建设，以缩小国家内部和国家之间的数码技术鸿沟。」为宗旨，于 2009 年 4 月在巴黎正式启用。该图书馆以阿拉伯文、中文、英文、法文、葡萄牙文、

¹⁴ <http://www.ulib.org>

¹⁵ <http://www.cadal.zju.edu.cn>

¹⁶ <http://www.worlddigitallibrary.org>

俄文和西班牙文向全球读者免费提供源于世界各地的包括手稿、地图、甲骨、善本、乐谱、录音资料、电影、印刷制品、照片、建筑制图等重要原始资料，可按时间、地点、主题和捐助机构等内容提供搜索和浏览服务。

世界数字图书馆是图书馆界的一个国际合作项目，2005 年由美国国会图书馆向联合国教科文组织提议建立，在其 1992 年发起的「世界记忆」工程基础上筹建，并得到一些公司和私人基金会（如谷歌公司、卡塔尔基金会、纽约卡内基公司、微软的公司、苹果公司、OCLC）的支持。参与创办的还包括中国、巴西、埃及、法国、伊拉克、以色列、日本、卡塔尔、俄罗斯、塞尔维亚等 28 个国家的图书馆。

虽然现在世界数字图书馆的资料数量还不是很多，但是质量都非常高，它们都是世界文化的精髓。WDL 代表着数字图书馆项目另一种风格：数量优先，质量第一。

欧洲数字图书馆（Europeana）¹⁷于 2008 年 11 月在布鲁塞尔正式开通，以文化艺术遗产为主，内容包括从欧洲各国博物馆、图书馆、档案馆和视听收藏中已经数字化的内容中选取的 200 万件电影、照片、绘画、声音、地图、手稿、图书、报纸和档案。

2004 年美国 Google 公布世界最大数字图书馆的建设计划后，法国总统希拉克发出「欧洲文化保卫战」的倡议。2005 年 4 月，由世界第二大图书馆法国国家图书馆牵头，欧洲 19 个国家图书馆发表联合共建欧洲数字图书馆的声明。同年，这一项目正式设立，并由欧盟负责协调和承担初期的资金投入。根据欧洲各国达成的协议，该数字图书馆的建设将分两个阶段进行。第一阶段的建设工作将主要围绕欧洲文化的奠基作品和欧盟各国的主要著作展开。第二阶段的工作全面铺开，欧洲各国的期刊、字典等各种图书都将完成数字化处理。

欧洲数字图书馆目前可以提供英语、德语、法语等 20 多种语言的服务。网站内容没有版权。

欧盟已投资 1.19 亿欧元，用于数字图书馆新科技的科研与开发。文学作品

¹⁷ <http://www.europeana.eu/portal/>

数字化开支用由各国自己承担。欧洲文化作品中只有约 1% 可以获得电子版本。计划 2010 年底前对 1,000 万册书籍进行数字化处理，欧盟委员会估计，成员国们大约还要总共投资 3.5 亿到 4 亿欧元。

谷歌数字图书馆（Google Books，早期亦称作「谷歌图书搜索 Google Print」）¹⁸是目前全球最为雄心勃勃的数字图书馆项目。谷歌创始人之一谢尔盖·布林（Sergey Brin）说：「Google 的使命是整合全球范围的信息，使人人皆可访问并从中受益。」谷歌的最终目标是创建人类最大的在线知识库，让人们可以自由使用，把世界各地的图书馆搬到任何能够上网的人们面前，而不再是只有在那些知名研究机构的研究人员才能享用。

该项目始于 2002 年，2004 年 10 月在法兰克福书展对外公布，初期与哈佛大学、密歇根大学、纽约公立图书馆、牛津大学和斯坦福大学合作，这些著名图书馆藏书总量预计超过 1,500 万卷。2006 年正式上线。目前已与世界各地超过 2,800 多家图书馆和 1 万多家出版机构与之建立了合作关系，已对大约 1,000 万册各种语言的图书做了数字化处理，能够实现 100 多万本书籍的在线全文搜索。全球用户均可免费通过图书检索部分或全文阅读图书，还能通过网上书店进行订购。

谷歌数字图书馆项目伊始，就因为知识产权等问题引发诸多国家和地区的出版界、IT 同行、各类团体和众多著作人的抗议和诉讼。在大陆更是搞的沸沸扬扬，至今余波未了¹⁹。

对谷歌数字图书馆的非议中最多的就是侵权问题。这是一个全球性的难题。在围绕纸质书籍而建立和完善起来的相关社会体系和法律环境中，知识产权问题的确是数字图书馆构建中一个难以逾越的门槛。人们一直在做着各种尝试和努力，迄今还没找到一种妥善之道，既可兼顾著作人和出版者的权益，同时又能满

¹⁸ <http://books.google.com/books?hl=zh-CN>

¹⁹ 2009 年 10 月，中国文学著作权协会发布谷歌「侵权」调查，570 位中国著作权人的 17,922 部作品在未取得了权利人授权的情况下被谷歌数字图书馆扫描上网。文著协之后与谷歌方面就这些作品的版权问题进行了三轮会谈，谷歌方面同意将赔偿标准从每本 60 美元提高到每本 300 美元。中国文学著作权协会 2009 年 12 月公布了谷歌公司提供的谷歌数字图书馆收录的中国图书初步清单。根据这份清单，谷歌数字图书馆共涉及中国图书 8 万余种，其中涉及中国作协会员约 2,600 人的约 8,000 种图书。中国著作权人同谷歌的谈判也因谷歌或退出中国市场的消息而暂停。

足数字时代大众便捷低成本获取纸质书籍内容的需求。相关利益与社会需求的平衡点还未找准。2008年微软宣布退出数字图书馆领域的原因就在于此。²⁰谷歌数字图书馆若不是一再强调并实践其公益性，恐怕也坚持不到今天。同样，大陆这些电子书资源项目都有各自的亮点和缺憾。如果说有什么共同点的话，那就是都尚未完全解决好知识产权问题。

在世界各地的数字图书馆饱受知识产权问题困扰艰难前行的同时，大量公益性免费电子书专业网站却层出不穷、百花齐放，已经积累了惊人数量的数字资源。纸书数字化并在网络上自由传播已然成为难以阻挡的世界性潮流。

纵观中文电子书资源领域十余年来的发展和现状，可以概括说：成绩不小、前景难料、愿景尚存。

电子书的出现使当代人类社会的信息环境发生了革命性的变化，目前还只是处在这场大变革的初级阶段。事实上，十年前就有消息称，美国人正在实施一项「最后一本书」的项目。在一部与笔记本电脑大小相仿的电子书籍中，除了将世界上最大的图书馆—美国国会图书馆馆藏所有信息囊括殆尽，还具有查询检索、增添新信息等功能，当时乐观估计这「最后一本书」将于两三年内面世。但时至今日实现这个目标的前景还是相当渺茫。无论是亚马逊的 **Kindle** 还是索尼的 **Reader Pro** 电子书阅读器现在可以阅读的书籍以达到几十万种。但仅仅做到方便阅读，很难做到信息的深度检索和知识挖掘。

尽管如此，人们依然「愿景尚存」，期待在不远的将来，通过互联网人们真正做到将人类文明有史以来的所有信息「一点即得」—这是 IT 界、信息界、知识界的共同努力目标。在这个过程中我们不应该只是纯粹的看客或者受益者，应当有所作为。

构建一个囊括古今中外所有汉语书籍的电子书库应是汉语圈人们的共同期待。这是一个必然要做的事情，问题在于谁来做，何时做，怎么做。两岸三院信息技术与应用交流研讨会已经举办了八届，在不同场合大家多次表达希望利用交流研讨会这个平台展开一些具体的合作项目的意愿。可不可以在这个领域做一些

²⁰ 微软已经扫描了 75 万册书籍和对 8,000 万篇期刊文章进行了索引。

尝试？譬如，由两岸三院牵头，联合和动员方方面面的力量，整合构建一个全球最大的中文电子书数据库。中研院在数字典藏方面已经做了大量卓有成效的工作，中科院在 e-Science、国家科技数字图书馆及全球数字图书馆等项目中完成很多基础性研究和实践，社科院在人文资源整合方面也做了一些有益的探索。以目前的各种条件来考量，由两岸三院牵头来做这件事应该是一个合适的选择，功在当代，利在千秋。

【作者简介】

王小宽 男

职 称： 社科院 计算机网络中心 副研究员

研究领域： 历史学、计量史学、图书馆、人文社会科学信息化

个人简介： 曾从事历史学、计量史学、图书馆、人文社会科学信息化等方面研究工作

联络电话： 86-13701257092

联络邮箱： wangxk@cass.org.cn