

·专稿·

# 数字期刊阅读量统计路径及方法初探

陈 丹 杨 萌

(北京印刷学院新闻出版学院 北京 102600)

**摘 要:**网络媒介的发展促进了网络阅读的发展,数字阅读作为一种新型的阅读形式,其阅读量的统计方式和方法还未形成标准。文章通过梳理国内外数字产品使用统计指标和方法,以及国内外现有数字期刊阅读量统计方法,从数字期刊有效浏览量的角度,以龙源期刊网为例,提出了关于数字期刊阅读量的统计路径及计算方法的构想。

**关键词:**数字期刊 阅读量 统计路径 统计方法

中图分类号: G255.2

文献标识码: A

文章编号: 1003-6938(2012)02-0018-05

## Exploring the Statistical Path and Method of the Digital Journal's Reading Quantity

**Abstract** The development of network medium has promoted the development of online reading, as a new way of reading, the statistical method of digital journal's reading quantity demands forming a standard. This paper taking qikan.com.cn for an example, using the statistical index and method of the digital products both domestic and abroad for reference, and also referring to the existing statistical method of the digital journal's reading quantity, to put forward the conception of digital journal's reading quantity and its calculation method from the angle of efficient page views.

**Keywords** digital journal ; reading quantity ; statistical path ; statistic method

从20世纪90年代中期开始,个人电脑和互联网在我国民众中迅速普及。据中国互联网信息中心(CNNIC)统计,截止到2011年12月底,中国网民规模达到4.57亿<sup>[1]</sup>。这不仅仅体现了信息技术的发展,实际上,计算机网络应用的大众化也正在悄然改变着人们获取信息、学习知识的途径,也深刻影响着人们的阅读习惯和思维方式。同时,网络的蓬勃发展所形成的新的读者群和内容生态环境,给传统期刊带来了巨大的发展空间。根据中国新闻出版研究院发布的《2010-2011中国数字出版年度报告》<sup>[2]</sup>数据显示,2010年中国数字出版产业总体收入为1051.79亿,其中数字期刊7.49亿,这说明数字期刊和期刊数字化阅读已为广大读者所接受。

在这一大背景下,出现了多家颇具规模的专业期刊网络传播平台和众多的个性化期刊数字化传播网站,许多期刊在进行传统出版的同时,也实现了数字化传播发行,一个新的内容产业已经形成。但是目前为止,在这个领域,还缺乏对数字期刊的发行统计

规则,这对行业转型的把控、产业的估量、著作人的版权收益统计、内容编辑加工质量评估、传播的人群对象分析等一系列问题显得缺乏依据,已阻碍了中国数字出版发展的进程,因此需要尽快制定一个统一的标准,对数字期刊的阅读量进行科学的统计分析,以此促进各方的利益分成,以科学发展观推进数字出版。

### 1 目前国内外数字出版物使用/浏览统计指标和方法

#### 1.1 网页浏览量的主要统计指标和方法

新闻出版总署2010年发布的《关于加快我国数字出版产业发展的若干意见》(新出政发[2010]7号)<sup>[3]</sup>文件提出:数字出版是指利用数字技术进行内容编辑加工,并通过网络传播数字内容产品的一种新型出版方式,其主要特征为内容生产数字化、管理过程数字化、产品形态数字化和传播渠道网络化。只要使用二进制技术手段对出版的各个环节进行操

作,应该都属于数字出版的范畴。因此,数字出版是传统出版业在内容和形式上的延伸和扩展,不仅大大丰富了出版物的内容和形式,也改变了传统出版物的生产方式和消费理念,直接创造出一些新的出版媒体形式。

作为开展数字出版的主要平台,各种类型的网站成为目前较为典型的数字出版物。目前,国际上提供网络访问数据分析和统计的网站主要有AC尼尔森、Google analytics、Alexa。从这些网站的排名机制中,可以看出它们并不是单单以网页浏览量(Pageviews)为排名依据,还会考虑到UB(Unique Brower)独立访问者、TS(Total sessions)全部会话即网站访问次数、ASD(Average session duration)平均停留时间、新访问次数百分比、跳出率、Search%来自搜索引擎的用户比例等等。所以,网站排名统计方法是多元性、多渠道的,这对数字期刊阅读量的统计具有一定的参考价值。

## 1.2 图书馆界对各类电子服务使用情况的统计

《国际图书馆统计标准 ISO2789》(以下简称《标准》)是国际标准化组织(International Organization for Standardization)与国际图书馆协会联合会(International Federation of Library Associations)为图书馆统计制订的国际标准,该标准既规定了传统图书馆服务的统计测量定义,又从服务类型、使用统计、附加调查方法三方面对电子图书馆服务使用情况提供了标准的统计测量定义<sup>[4]</sup>。

为了及时了解和掌握图书馆提供的各类电子服务使用情况究竟如何,《标准》从图书馆提供电子服务的利用形式、使用地点和对电子服务的分项数据收集等三方面提供了被认为最能反映图书馆电子服务使用情况并需要测量的数据项。在利用形式方面,《标准》确定了四个核心数据组:任务数(Number of Sessions)、文献下载数(Number of Documents Downloaded)、检索记录下载数(Number of Records Downloaded)和虚拟访问数(Number of Virtual Visits)。除了这些核心数据组提供电子服务使用基本信息外,图书馆在可能的情况下也应收集一些相关附加的数据。这些附加数据是:任务检索时间(Session Time)、被拒绝访问任务数(Number of Rejected Sessions)、检索次数(Number of Searches)、通过因特网访问的任

务数(Number of Internet Sessions)。上述核心数据项与附加数据项较为清晰地反映出图书馆电子服务不同利用形式的情况,也是图书馆电子服务使用统计的主要内容。另外,《标准》还提出图书馆电子服务要区分三种使用地点。即(1)在图书馆内部使用;(2)虽在图书馆外部,但却在机构内部或授权范围内使用;(3)在机构或授权范围之外使用。通过电子服务三种使用地点的统计数据,能够进一步了解图书馆电子服务多大程度上扩大了图书馆服务的范围,在多大程度上增加了市场辐射力。第三,对于各项服务,如OPAC、电子馆藏、网站等,由于利用形式和测量方式的不同,还应该分类收集各服务形式的数据。主要包括OPAC数据、电子馆藏数据、图书馆网站访问数据等。

对于图书馆所提供的电子服务,我国图书馆统计标准为GB/T13191-2009,《标准》建议收集以下三种核心数据集:会话次数、下载的次数(细分为内容下载和记录下载)、虚拟访问的次数。除了这些提供电子服务利用基本信息的核心数据集之外,《标准》建议尽可能采用其他一些数据收集:会话时间、被拒会话(受阻进入)次数、检索次数、因特网会话次数。可见,以上所有的数据都是对图书馆馆藏、图书馆网站、联机目录和经由图书馆提供的因特网访问利用的统计,而不统计用户在图书馆内登录因特网获取的公开免费的网络文献。

## 2 数字期刊阅读量及其统计思路的提出

### 2.1 数字期刊阅读量的提出

阅读是从书面材料中获取信息的过程,是读者和文本的对话。阅读是一种主动的过程,是由阅读者根据不同的目的加以调节控制的。对于传统期刊的阅读,无论是发行者,还是广告商,都主要以该期刊的发行量为考量指标。同理,对于数字期刊的阅读,主要是考量阅读主体(读者)与数字化文本相互交流信息与知识的程度,也即数字期刊中数字内容被网民获取的信息量多少。通过前面对网页浏览量以及图书馆各类电子服务使用情况的统计,可以看出,对于数字内容而言,浏览量与阅读量还是有所不同的。由于浏览包括主动、被动浏览(如自动弹出网页等),浏览时读者对呈现内容的接受程度也有所不同,因

此,本文建议将“有效浏览量”与“阅读量”等同起来,如在读者进行数字期刊浏览时,只有进行有效的浏览,才可以作为数字期刊阅读量进行统计。

为了做到数据的科学性和全面性,在数字期刊阅读量这一计算体系中,核心是“付费点击”,也就是说,只要是用户通过付费模式的点击都可以成为数字期刊阅读量的有效组成部分。另外,点击浏览量,包括搜索引擎抓取数字期刊服务提供商的相关页面的点击次数(是指网络用户通过搜索引擎而找到的相关杂志付费页面),也应该成为期刊网络阅读量的有效组成部分。

## 2.2 国内外数字期刊阅读量统计方法概览

在MPA美国杂志发行公会的统计中,期刊的网络阅读量主要是指通过网络发行的原貌版电子期刊的总量。苹果公司是在2010年4月发布iPad的,而美国发行调查局(Audit Bureau of Circulations,简称ABC)在iPad发布前,就意识到了iPad将给数字报刊领域带来巨大的影响。于是,在2010年的3月15日,ABC就提前修改了自己对“数字报刊阅读量”的定义,以适应iPad为数字报刊行业带来的变化。ABC的规定是:只要包含印刷版的所有内容和广告,付费数字报刊用户就可以计入报刊阅读量。而之前的规则是,数字报刊的内容只有与印刷版的报刊完全相同,才能被计入阅读量。

根据ABC的规则,统计数字报刊阅读量有两个必要条件:内容(包括文字和图片)和广告。之前,虽然有一些报刊可以在亚马逊Kindle上阅读,但是并不被计入到它们的有效阅读量中,原因就是它们不具备广告。纸版杂志上的小广告,如分类广告,或者是不超过版面1/3大小的广告,可以不出现在杂志的数字版上。只要保证电子版上必须出现纸质版的广告,那么数字版杂志同样可以刊登纸版杂志上没有的广告。

网络电子资源在线使用统计标准COUNTER的开发源于英国的PALS。2002年3月,在AAP、ALPSP,ARL,ASA,BIC/ED IT EUR,JISC,NCLIS,NISO,PA,STM等专业研究机构以及Blackwell Publishing,EBSCO,Elsevier Science,ProQuest,Oxford University Press等国际著名出版商的支持下,正式开展网络电子资源在线使用统计标准COUNTER的研究。2002

年12月,发布了COUNTER工作规范的第一个版本,该版本主要用于期刊和数据库在线使用统计信息的记录和交换。其工作规范的主要内容包括:定义用户计量的术语和数据元素;定义期刊和数据库使用报告的内容和格式;定义数据处理和审计要求;明确使用的指导原则等。其用户对包括图书馆、出版商、联机销售商及其它中介服务机构。COUNTER工作范围建立在现有的一些研究倡议和标准的基础之上,主要有:ARL NEW Measures Initiative、ICOLC Guidelines for Statistical Measures of usage of Web-based Information Resources、NISO Forum on Performance Measures and Statistics for Libraries and NISO Standard Z39.7<sup>[5]</sup>。

## 2.3 国内数字期刊阅读量统计方法

在国内从事期刊数字出版的行业有中国知网、万方、维普、龙源、悦读、读览等企业,作为国内数字期刊的领跑者之一,龙源期刊网积极参与、促进期刊的网络传播,在多年的发展中,龙源从2005年开始每年10月份对过去一年各刊的访问进行统计分析,并做公开发布,从而形成了龙源期刊TOP100文章排行榜、TOP100期刊排行榜等数据。同时,龙源还依靠强大的技术支持,为3000多种合作期刊定期提供详细的内容传播数据。这些数据均来自于计算机网络系统对用户阅读行为自动进行的统计,科学、客观、真实。

龙源期刊网在向社会开放的内容具有丰富多彩的呈现方式,可以大大丰富读者的阅读体验。比如,原文原貌版,保留了杂志原有版式,用户可以翻页阅读,仿佛一册在手。另外还有文本版,文本版则将杂志中的文字还原到网页上,可以复制、粘贴,保留每一篇文章和重要的图标和图片,最适合边阅读、边收藏的网络读者。人声语音版充分利用网络多媒体的特色,聘请电台专业的播音员把优秀的杂志内容变成可用手机、MP3、MP4下载的有声杂志,可以向无暇上网的人以及盲人提供服务。此外,还有手机版和PAD版阅读。其中,用户在阅读原文原貌版时,可以进行浏览,也可以放大阅读。在龙源期刊网的显示数据中,放大阅读量即为用户付费整刊阅读量,浏览量是用户点击浏览杂志的数量(除封面外用户只能放大浏览四页)。用户在阅读文本版时,有文本全文阅读和片段阅读两种模式,与原貌版阅读类似,全文阅

读为付费阅读,而片段阅读是用户对文章的一部分进行浏览(在龙源期刊网中,片段浏览字数是全文首字起到300字之后的第一个句号止)。另外手机版阅读则是指对期刊的付费阅读。

其中,用户在龙源期刊网上付费点击的期刊次数包括对整本刊物(原貌版)的付费点击次数、对期刊中单篇文章(文章版)的付费点击次数、以及对其他类型(语音版、手机版等不同产品版本)的付费点击次数之和,这些都是基于服务器的技术统计数据经过精确的计算而生成的;对于读者的点击浏览,龙源认为,网络阅读量不能沿袭传统纸媒的算法——卖掉一本算一本,网络阅读过程中不管读者有没有购买,他封面看到了、目录看到了、甚至每篇文章的开头看到了,虽然没有看到全文,但是达到了一定传播效果,这个也应该算作一种阅读,与付费阅读的阅读量若作区别的话,可称之为次有效阅读量。

故用户在龙源期刊的阅读量由两部分组成:一是有效阅读量,即用户在期刊网上付费点击的期刊次数;另一种是次有效阅读量,即读者点击浏览次数,其中包括读者通过搜索引擎抓取的龙源相关页面从而进行点击的次数。其中,用户在龙源期刊网上付费阅读的期刊次数包括对整本刊物(原貌版)的付费点击次数、对期刊中某篇文章(文章版)的付费点击次数、以及对其他类型(语音版、手机版等不同产品版本)的付费点击次数。而用户在龙源期刊网上点击浏览的期刊次数包括对整本刊物(原貌版)的浏览,对期刊中某篇文章(文章版)的片段阅读次数。龙源在统计其期刊的阅读量时,会根据读者浏览的“有效性”,综合考虑这两种阅读量数据。

总之,以上的几种统计方法,对我们开展数字期刊阅读量的统计具有一定的借鉴意义。但是,对于数字期刊浏览量以及阅读量的统计和评价,不管是龙源期刊网还是相关业界目前并没有较为科学和权威的标准。数字化内容的浏览与阅读量之间有何关系,如何较为科学地统计数字期刊的阅读情况?本文,将以龙源期刊网为例,对数字期刊阅读量统计方法提出构想。

### 3 数字期刊阅读量统计方法的构想

#### 3.1 数字期刊阅读量统计要素组成

根据对龙源期刊网数字阅读数据分析可以看出,不管是文本版阅读还是原貌版阅读,有很大一部分用户会选择片段阅读、期刊浏览。根据前面我们对数字期刊阅读量概念的界定,数字期刊阅读量主要是考量阅读主体(读者)与数字化文本相互交流信息与知识的程度,也即数字期刊中数字内容被网民获取的信息量多少。因此,除了全文阅读文本版期刊文章、整本下载阅读原貌版文章以及手机阅读文章的数量外,读者对期刊文本片段的阅读量以及对期刊原貌版的浏览量应该也成为数字化产品阅读量统计的有效组成部分。

通过前面的论述可知,目前国内外对于如何统计数字内容产品(如网页)的浏览量,已有一些较为公认的方法和指标,那么,如何对浏览的“有效”性进行考量,即对本文提出的“数字期刊阅读量”进行统计,是本文要研究的重点。笔者认为,数字期刊阅读量的统计,应以人们阅读数字期刊中的文章数为基本单位,包括对整本刊物(原貌版)的付费点击次数、对期刊中某篇文章(文章版)的付费点击次数、对其他类型(语音版、手机版等不同产品版本)的付费点击次数以及用户对期刊中某篇文章(文章版)的片段阅读次数、对整本刊物(原貌版)的浏览次数等各种要素。

#### 3.2 数字期刊阅读量统计公式

前面提到,龙源期刊网目前统计数字化产品阅读量的方式是以付费与否为取舍标准的,这从TOP排行的角度,有一定的道理。但作为数字期刊的阅读量统计方式,有必要提出一种更为全面的统计方法来计算数字期刊的阅读量。

前面,我们已经提出数字期刊阅读量的概念。这里将对期刊阅读量的计算方法给予说明。

由于数字期刊阅读量以文章数为基本单位,对于龙源期刊网相关的换算,可以分为两类,一类是关于原貌版的换算,包括对整本刊物(原貌版)的付费点击次数、整本刊物(原貌版)的浏览次数进行换算,另一类是对文本版进行换算,即对某篇文章(文章版)的片段阅读次数进行换算。首先,对原貌版进行换算:

(1) 对整本刊物(原貌版)的付费点击次数的换算。一本期刊一般会有40-50篇文章,但针对某一本



期刊,文章数各不相同,用系数 $\lambda$ 表示。

(2)对整本刊物(原貌版)的浏览次数进行换算。龙源期刊网上对于原貌版整本刊物的放大阅读除封面外只能有四页,把其换算成期刊数,应为 $\frac{4}{\text{期刊总页数}}$ 。再根据我们对于原貌版整本刊物的换算方式,把原貌版浏览次数换算成数字期刊阅读量应该是 $\frac{4\lambda}{\text{期刊总页数}}$ 。那么,对于一般的原貌版期刊,其数字期刊阅读量是 $\frac{\text{试读页数}\lambda}{\text{期刊总页数}}$ 。

(3)对某篇文章(文章版)的片段阅读次数的换算。龙源期刊网上的片段浏览字数是全文首字起到300字之后的第一个句号止,根据这一情况,取一个中间值350字。针对每一篇不同的文章,字数各不相同,所以用变系数来表示某篇文章片段阅读次数的换算: $\frac{350}{\text{文章总字数}}$ 。那么,对于一般的文章版期刊,其片段阅读次数是 $\frac{\text{试读字数}}{\text{文章总字数}}$ 。

另外,由于语音版的数据非常少,不将其作为计算数字期刊阅读量的组成成分。

那么,整本刊物(原貌版)的付费点击阅读与整本刊物(原貌版)的浏览次数之和换算成阅读文章数即为原貌版的有效阅读量,期刊中某篇文章(文章版)的付费点击次数与期刊中某篇文章(文章版)的浏览次数之和即为文本版的有效阅读量,二者之和再加上手机版的付费点击次数即为期刊的有效阅读量。

用字母表示数字期刊阅读量的计算方式,其中,  
Q:数字期刊阅读量

A:期刊中某篇文章(文章版)的付费点击次数

B:期刊中某篇文章(文章版)的片段阅读次数

C:整本刊物(原貌版)的付费点击次数

D:整本刊物(原貌版)的浏览次数

E:其他类型(语音版、手机版等不同版本)产品的付费点击次数

$\lambda$ :对整本刊物(原貌版)的付费点击次数的换算系数

由此得出数字期刊阅读量的计算方式为:

$$Q=A+\frac{\text{试读字数}}{\text{文章总字数}}B+\lambda\left(C+\frac{\text{试读页数}}{\text{期刊总页数}}D\right)+E$$

## 4 结语

本文以龙源期刊网为例,对其数据进行分析,从数字期刊的有效浏览量的角度,提出了数字期刊的阅读量的概念以及计算方法。这对于制订数字期刊阅读量的统计标准,具有一定的借鉴意义。但由于数字阅读是一个涉及读者分布的区域性、读者阅读的目的性、原刊品牌影响力、期刊的产品形态、网络推介方式以及期刊收费情况等因素的复杂系统,因此本文对其阅读量的统计只涉及了读者对期刊文章的直接点击情况,对数字内容提供商提供服务的类型和方式,以及读者对数字期刊的利用形式、阅读的地点以及阅读路径等因素及其影响还未涉及。这也为今后数字期刊及其他数字出版物阅读量的统计路径及方法的完善提供了新的思路和视角。

总之,标准化数字期刊阅读量的统计方法,不仅可以扩大期刊营收的市场份额,同时也是期刊业走数字化出版之路的必然要求,它对于期刊社在互联网环境下吸引新生代读者和广告主,以及图书馆进行数字内容服务使用情况统计、用户分析和开展相关研究等都颇有意义。

参考文献:

- [1]CNNIC发布《第29次中国互联网络发展状况调查报告》[EB/OL].[2012-01-29].[http://www.cnnic.net.cn/dtygg/dtgg/201201/t20120116\\_23667.html](http://www.cnnic.net.cn/dtygg/dtgg/201201/t20120116_23667.html).
- [2]2010-2011中国数字出版年度报告[EB/OL].[2012-01-29].<http://wenku.baidu.com/view/96aa8b020202740be1e9b89.html>.
- [3]版署关于加快我国数字出版产业发展的若干意见[EB/OL].[2012-01-29].<http://www.techweb.com.cn/news/2010-09-16/683222.shtml>.
- [4]金明华.国际图书馆统计标准ISO2789:2003(E)电子服务使用统计评析[J].现代图书情报,2006,(1).
- [5]National Information Standards Organization [EB/OL].[2012-01-29].<http://www.niso.org/home/>.

作者简介:陈丹(1969-)北京印刷学院新闻出版学院教授,研究生导师,研究方向:数字出版,数字阅读;杨萌(1987-)北京印刷学院2011级传播学研究生。