



电子文献学引论

Studies on Electronic Bibliography



蔡先金 赵海丽 编著



电子工业出版社

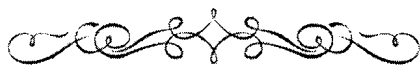
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

<http://www.phei.com.cn>



电子文献学引论

Studies on Electronic Bibliography



蔡先金 赵海丽 编著

电子工业出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

本书从传统文献学理论和电子信息技术知识入手,运用文献学的基本范畴和基本原理,对电子文献进行分析与描述、挖掘与探索,然后构建一个较为完善的电子文献学框架体系。本书共9章,主要内容包括:电子文献释义,电子文献的类型与编目,电子文献的生成、存储与传输,电子文献的文本与超文本,电子文献检索,电子文献的开发与利用,电子文献保护,电子文献评价等。

本书是文献学学术专著,也可作为高等学校出版发行学、图书情报学、古代文学、古代文献、教育学及其他文化哲学类专业相关课程的教材。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

图书在版编目(CIP)数据

电子文献学引论 / 蔡先金, 赵海丽编著. —北京: 电子工业出版社, 2012.8

ISBN 978-7-121-17274-8

I. ①电… II. ①蔡… ②赵… III. ①电子文献—研究 IV. ①G255.75

中国版本图书馆 CIP 数据核字(2012)第 119175 号

策划编辑: 王羽佳

责任编辑: 王羽佳 特约编辑: 曹剑锋

印 刷: 三河市双峰印刷装订有限公司

装 订:

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编: 100036

开 本: 720×1000 1/16 印张: 14.5 字数: 252.5 千字

印 次: 2012 年 8 月第 1 次印刷

印 数: 1000 册 定价: 45.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888。

质量投诉请发邮件至 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线: (010) 88258888。



前 言



电子文献是信息时代之产物,表现出不同于传统文献的许多特质,这使我们对电子文献进行研究成为了一种可能。从传统文献学理论和电子信息技术知识入手,运用文献学的基本范畴和基本原理,对电子文献进行分析与描述、挖掘与探索,然后构建一个较为完善的电子文献学框架体系,就成为了电子文献研究者的一项重要责任与任务。

以电子文献概念界定作为逻辑起点,指出电子文献具有两个最基本的特性:一是依托计算机、光盘、磁盘或其他电子形式的载体而存在;二是通过计算机技术记录、存储、传递、检索和浏览文献信息。基于此,再对传统文献与电子文献两者之间的区别与联系进行了必要的辨析。

从目录学与分类学原理出发,按照存储方式、载体形式、编辑与使用方式、加工层次、文献格式等不同分类标准与办法,将电子文献分为不同的类型,并对这些不同的电子文献类型予以梳理与分析。

以现代科学与技术发展的眼光,描述了电子文献的生成方式,包括录入、扫描、摄像、刻录、多媒体制作,探究了磁带、软盘、光盘、硬盘等存储介质的工作原理以及数字化网络出版等传输方式。

从文本学理论角度,研究电子文献具有的文本与超文本的形态,并探讨文本的自动摘要、超文本导航、文本的计量等操作性技术的运用。

从传统文献与电子文献比较入手,挖掘了计算机文献资源检索与网络文献检索的优势,并对于不同的搜索引擎进行了必要的功能价值比较,发现其优劣。

从实用价值层面,探究电子文献开发与利用的重要意义,介绍电子文献收集、知识获取、个人数字图书馆创建、文献资料处理的工作原理以及操作路径。同时,剖析电子文献使用中的非安全因素,寻找切实可行的保护措施。

从评价学的视角,构建了电子文献评价的指标体系,评价指标主要有原文

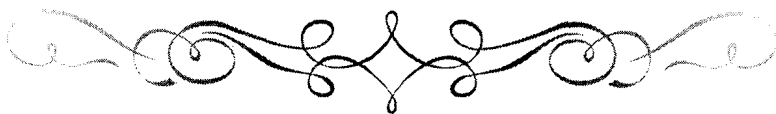


献质量、技术可靠性与运行稳定性、保密性与防病毒能力、检索多样性与便捷性、文献运行环境和网上运行速度以及页面设置状况等。

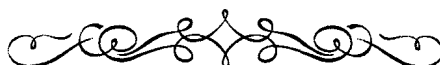
从文化哲学思考的角度，面对海量的现代电子文献，反思这个由“比特”构建的世界，质疑人在“创世纪”中的作为，追问人类是否应该“知止”而后生。

总之，电子文献学的研究是一个横跨多学科的课题，需要运用整体思维的方式去予以审视与研究，才有可能建立起一个全新的学科体系。

作 者



Preface



Electronic bibliography is born in the age of information, showing a number of characteristics being different from the traditional bibliography, which gives us a possibility of studying on electronic bibliography. From the theories of traditional bibliography and electronic information technology, a more comprehensive framework for electronic bibliography discipline should be built by the analysis and exploration of electronic bibliography.

TElectronic bibliography has the two basic characteristics: First, being on CD-ROM, disk or other electronic form; second, browsing information through computers by recording, storage, transmission and retrieval.

TElectronic bibliography in accordance with storage methods, carrier forms, editing and use patterns, information processing level and the different formats will arrive at different classification, which is conducive to recognize different electronic bibliography.

TElectronic bibliography is generated by entry, scanning, photography, recording, multimedia production, etc. Its storage mediums would be a tape, a floppy disk, CD-ROMs, hard drives. Its transmission ways would be a digital network publishing, etc.

Electronic bibliography forms can be divided into text and hypertext, which can be automatic summaries, hypertext navigation and the measure of the text.

TElectronic bibliography search ways can be divided into the computer resources search and digital bibliography retrieval, which search engines have many types with different features.

TDevelopment and use of electronic bibliographies have a great significance,



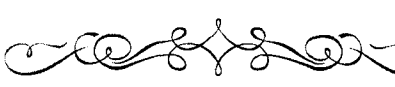
including the collection of electronic bibliographies, knowledge acquisition, creating personal digital libraries, documentation processing. At the same time, we should establish a sense of security for electronic bibliographies and implement practical measures of protection to prevent non-security factors.

Evaluation of electronic bibliography has a different item system to evaluate the quality of the original literature, technical reliability and operational stability, confidentiality and anti-virus capabilities, diversity and ease of retrieval, documentation, operating environment and web-speed as well as the homepage designing style.

Finally, facing a deluge of modern electronic bibliographies, we should reflect on the cultural philosophy and ask any questions, which include reflecting on the world built by the “bits”, questioning people who would act as the “Genesis” Lord, asking whether people should know the significance of rhythms and then know how to live.



目 录



第 1 章 导论	1
1.1 电子文献释义	2
1.1.1 电子文献概念	3
1.1.2 电子文献的特征	5
1.2 电子文献与纸质文献之比较	7
1.2.1 两种文献的交互方式比较	8
1.2.2 两种文献的形态比较	11
1.2.3 两种文献利用方式比较	13
1.3 电子文献产生的背景及发展过程	14
1.3.1 电子文献产生的背景	15
1.3.2 电子文献的发展过程	19
1.4 电子文献学的研究方法	22
1.4.1 电子文献学的提出	22
1.4.2 研究的理论基础	23
1.4.3 传统文献学研究方法与现代科学研究方法相结合	24
本章参考文献	25
第 2 章 电子文献的类型与编目	26
2.1 电子文献的不同分类法	28
2.1.1 按照结构形式分类	28
2.1.2 按载体形态分类	29
2.1.3 按照传媒体系分类	29
2.1.4 按信息加工层次分类	32



2.1.5 按文献格式分类	32
2.2 类型文献及其载体的发展进程	36
2.2.1 电子图书的发展进程	36
2.2.2 电子期刊的发展过程	38
2.2.3 数据库的发展过程	39
2.2.4 光盘的发展过程	42
2.2.5 半导体和磁泡存储器	43
2.2.6 其他光学存储载体	44
2.3 电子文献编目	44
2.3.1 机读目录	45
2.3.2 计算机编目系统	46
2.3.3 电子文献的编目方法	48
本章参考文献	49
第3章 电子文献的生成、存储与传输	50
3.1 电子文献的生成	51
3.1.1 录入	51
3.1.2 扫描	52
3.1.3 摄像	53
3.1.4 刻录	53
3.1.5 多媒体	54
3.2 电子文献的存储	54
3.2.1 信息存储技术简介	54
3.2.2 文献管理系统的结构与技术	58
3.2.3 电子文献的储存特点	64
3.2.4 电子文献的储存介质	65
3.2.5 电子文献的储存技术	71
3.2.6 国内重要数据库	74
3.3 电子文献的传输	85
3.3.1 接入因特网 (Internet)	85
3.3.2 远程登录 (Telnet)	85
3.3.3 FTP 文献传输	86



3.3.4 电子邮件 (E-mail) 的传输	87
3.3.5 BBS 的公告	88
3.3.6 数字化网络出版	90
3.3.7 博客	92
本章参考文献	94
第 4 章 电子文献的文本与超文本	96
4.1 “文本”释义	97
4.2 文本与版本之关系	98
4.2.1 版本与文本比较	98
4.2.2 古籍电子文献的版本	100
4.3 何谓“超文本”	102
4.3.1 超文本概念	102
4.3.2 超文本要素与特点	103
4.4 超文本的发展进程	106
4.5 文本与超文本的使用	110
4.5.1 文本的自动摘要	110
4.5.2 超文本导航	111
4.5.3 文本的计量	112
本章参考文献	113
第 5 章 电子文献检索	114
5.1 电子文献检索发展历程	116
5.1.1 脱机检索阶段	116
5.1.2 联机检索阶段	117
5.1.3 光盘检索阶段	118
5.1.4 网络检索阶段	119
5.2 网络检索工具	120
5.2.1 网络检索工具的类型	120
5.2.2 搜索引擎	124
5.2.3 搜索引擎评价	128
5.2.4 若干 Internet 资源搜索引擎简介	129
5.3 引文索引	134



5.3.1 引文索引概说	134
5.3.2 国内外重要引文索引	135
5.3.3 引文分析法	139
本章参考文献	141
第6章 电子文献的开发与利用	142
6.1 电子文献开发之意义	143
6.1.1 可以推动社会文明进步	143
6.1.2 可以促进自然资源的开发利用	143
6.1.3 可以挖掘与展示文献的社会功能	144
6.1.4 可以为社会实践活动提供情报和信息	144
6.2 电子文献收集	144
6.2.1 电子图书的收集	144
6.2.2 电子期刊、报纸的收集	150
6.3 知识获取	151
6.3.1 数据挖掘	152
6.3.2 Web 挖掘	152
6.3.3 文本挖掘	156
6.4 创建个人数字图书馆	157
6.4.1 个人数字图书馆的概念	157
6.4.2 个人数字图书馆的功能	158
6.4.3 个人数字图书馆的构建	159
6.5 文献资料的处理	160
6.6 电子文献利用率	162
6.6.1 影响电子文献利用率的主要因素	162
6.6.2 提高电子文献资源利用率的对策	163
本章参考文献	164
第7章 电子文献保护	165
7.1 电子文献的非安全因素	166
7.1.1 电子文献的非安全因素	166
7.1.2 电子文献损失与破坏的原因	168
7.2 电子文献保护措施	172



7.2.1 电子文献保护项目	172
7.2.2 电子文献安全规范	173
7.2.3 电子文献的物理保护	173
7.2.4 电子文献的逻辑保护	179
7.2.5 电子文献的内容保护	181
7.2.6 电子文献的版权保护	188
7.2.7 病毒防治	189
7.2.8 “黑客”防范	194
7.3 电子文献的安全意识	196
本章参考文献	197
第 8 章 电子文献评价	198
8.1 评价指标及体系特征	199
8.1.1 评价指标	199
8.1.2 评价体系的特征	201
8.2 评价主要内容	202
8.2.1 原文献质量	202
8.2.2 技术性评价	204
8.2.3 保密性与防病毒能力评价	205
8.2.4 检索多样性与便捷性评价	206
本章参考文献	207
第 9 章 结语	208
9.1 是粒子的世界？还是比特的世界？	209
9.2 是神在“创世纪”？还是人在“创世纪”？	211
9.3 是不知停歇？还是“知止”？	212
本章参考文献	213
参考文献	215
后记	219

第 1 章
导 论



文献起源之早,可谓与人类知识萌生同步,而后在人类文明发展进程之中,乃构成人类全部知识学重要基础条件之一。^①相对于当下的海量文献而言,文献既可泛化为载体上人类知识谱系之通称,又可谓载体上知识纲属门类之特称,造成如此称谓之主因是缘于信息时代电子文献新种类的出现。自从电子文献实现网络形式传播之后,可以说,文献充斥了整个虚拟网络世界,以至于人们无暇顾及文献一词的古典含义,更不会去作词源考索之工夫,但这并不意味着否定该词最早见于传世之《论语·八佾》之定说^②,亦不可以推翻宋末元初马端临在《文献通考·自叙》中对该词之解释。^③只是随着信息时代之来临,文献认识、文献处理和文献传播方式在新的科学技术推动下发生了革命性的变化,电子文献仅凭借其特有的品格以及神奇的“魔法”,正在开辟文献学的一个崭新的时代,此乃集中体现了我们这个时代与时俱进之精神。

1.1 电子文献释义

电子文献是信息时代的产物,表现出不同于传统文献的许多特质,这使我们对电子文献进行研究成为一种可能。黑格尔认为:“就对象来说,每门科学一开始就要研究两个问题:第一,这个对象是存在的;其次,这个对象究竟是什么。”^[1]电子文献作为研究对象之存在是勿需置疑的,现在只是需要回答:这个对象究竟是什么,也就是说,需要为电子文献下一个正确的定义,当然要建立在文献之古典定义的基础之上,因为任何事物的产生与发展都是渊源有自的。

① 此处引用“知识学”之概念,不应混同于德国费希特将自己的哲学称谓为“知识学”之概念,从广义的文献来说,此处的“知识学”应该是“文献学”之放大。^[2]

② 《论语·八佾》:“子曰:‘夏礼,吾能言之,杞不足徵也。殷礼,吾能言之,宋不足徵也。文献不足故也。足,则吾能徵之也。’”

③ 元代马端临始用“文献”名其书,即《文献通考》。张舜徽在《中国文献学》中云:“用‘文献’二字自名其著述,起于宋末明初的马端临。他写了一部贯通历代典章制度的《文献通考》。《自叙》指出:‘凡叙事,则本之经史而参之以历代会要,以及百家传记之书,信而有证者从之,乖异传疑者不录,所谓文也。凡论事,则先取当时臣僚之奏疏,次及近代诸儒之评论,以至名流之燕谈,稗官之记录,凡一话一言,可以订典故之得失,证史传之是非者,则采而录之,所谓献也。’”^[3]



1.1.1 电子文献概念

在传统古典文献学领域,由于文献的载体如纸张、竹简、甲骨等表现出的物质的有形性特点为世人所熟知,前人在定义文献这一概念时便更加关注于其内涵的界定,即指向典籍的记载与宿贤的言论,这亦是词源考稽的结果。^①是故,中国古典文献学开山之人郑鹤声、郑鹤春认为:“结集、翻译、编纂诸端,谓之文;审订、讲习、印刻诸端谓之献。叙而述之,故曰文献学。”^[4]传统的文献定义虽然有其存在的道理,但是随着时代的发展亦表现出其在所难免的局限性。而今“文献”之定义显露出与时俱进的时代精神,可从国家标准的制度层面予以较为科学的确认。1983年我国颁布的《中华人民共和国国家标准·文献著录总则》(GB3792.1—83)给“文献”下了一个简明的定义是:“记录有知识的一切载体。”而国际标准化组织《文献情报术语国际标准》(ISO/DIS5217)对“文献”的定义是:“在存储、检索、利用或传递记录信息的过程中,可作为一个单元处理的,在载体内、载体上或依附载体而存储有信息或数据的载体。”实质上,国际化标准组织对于文献的定义与我国的国家标准定义之间并不矛盾。无论如何,文献之定义不能缺少两个最基本的构成要素:一是要有一定的知识或信息内容,没有记录任何知识或信息内容的纸张、磁带等载体不能称为文献;二是要有用以记录知识或信息的物质载体,没有载体的知识亦不能称为文献,从这个意义上说,存在于人们头脑中的知识或信息不能称之为文献。因此,我们可以给出这样的定义:文献是指那些将知识、信息用文字、图像、数码等各种符号,通过书写、印刷或其他诸如光学、电磁学等方法记录在一定物质载体上的结合体。^②

伴随着多媒体、互联网、电子邮件、电子图书馆、数字化信息等一系列让

① 许慎《说文解字·叙》云:“仓颉之初作书,盖依类象形,故谓之文。”郑玄《论语》注云:“献,犹贤也。”^[5]朱熹《论语集注》云:“文,典籍也;献,贤也。”^[6]张舜徽《中国文献学》云:“过去封建学者们所强调的‘征文考献’,便是说要了解过去的历史,一方面取证于书本记载,一方面探索于耆旧言论。”

② 信息与知识两个概念之间既有区别也有联系。信息是指主体通过各种认知方法获得并经过收集、记录、处理后以某种形式存储起来的事实或数据,具有物质性、传递性、可扩充性、共享性、价值不定性等特性。知识是人类社会实践的总结,是人的主观世界对客观世界的如实反映,包括通过阅读、学习等方式的认知和通过自己或他人的历练所积累的经验的总和。知识同信息的关系就是将反映自然现象和社会现象的信息经过加工,上升为对自然和社会发展客观规律的认识。



人耳目一新的概念的出现,人们越来越多地关注电子文献这一概念。人们从不同的角度,围绕电子文献的主题开展了许多有益的讨论,并开列出了一些不同之称谓,如电子图书、电子期刊、电子报纸、电子出版物、电子资源、电子资讯、网上出版物、网上文献等。显然,认识不同,称谓则不一,并对研究、交流和传播电子文献造成一些障碍和不便。因此,给电子文献以科学、合理的界定就显得非常必要。1997年11月,国际标准化组织在国际电工委员会的协助下颁布了国际标准ISO690—2:1997(E)。该标准分别就电子文献及其相关定义进行了简要阐述和说明,它所列举的18个术语概念多数是在原有国际标准的基础上修订、改进和补充而来的,而像公告板、数据库、讨论表、电子文献等概念则是第一次在ISO中出现,将电子文献概念简洁而明了地界定为:以电子形式存在的通过计算机技术存取的文献。^①由此我们可以给电子文献概念勾勒出如下一个架构:一是电子文献是某种特定的文献;二是它要依托计算机、光盘或其他电子形式的载体存在;三是它要借助计算机技术记录、存储、传递、检索和浏览信息。因此,电子文献是应用计算机技术、现代通信技术、高密度存储技术和网络技术等高新技术开发的新一代文献,并以电子形式(包括数字代码等),将文字、图、声、像等信息存储在磁、光、电介质上,通过计算机或类似设备阅读使用。

电子文献构成需要具备这样两个基本要素:一是“以光盘、磁盘或其他电子形式存在”;二是“通过计算机技术记录、储存、传递、检索和浏览信息”。这也是电子文献所具有的两个最基本的特性。换言之,“两要素”或“两特性”决定了电子文献与其他文献之不同。如果按照这一原则来确定电子文献范畴,那么,当电子计算机出现的同时,专供计算机识读的计算机磁带就是一种电子文献,而可供计算机识读的纸带式文献由于本身材料是以纸质载体形式而不是以电子形式存在的,因而不属于电子文献范畴。值得注意的是,伴随着电子计算机的出现,计算机文件(file)也同时出现。该“文件”是指按一定顺序存放的、相互联系的一个完整的、有名称的信息的集合,如程序、程序所用的一组数据或用户创建的文档。文件是基本的存储单元,它使计算机能够把一组信息与另一组信息区分开来,用户可以对其进行检索、修改、删除、存储或发送到输出设备。^[7]按照它的这一定义,计算机文件也同时满足电子文献构成的两

① 蒙戴尔(Craig Mundiel)则将电子文献定义采用描述性的语言叙述为:“以计算机格式获得的,通过发表在中心数据库出版的,能以软盘、CD-ROM 互通信连接传播的,并可用计算机快速浏览的文献。”



个基本要素或基本特性,因此,毫无疑问,计算机文件属电子文献范畴。但是,1912年出现的缩微平片、1915年以后出现的目视重叠比孔卡、1940年代出现的穿孔卡片,以及后来出现的缩微胶片等文献,由于它们本身不是以计算机技术记录、储存、传递、检索和浏览信息的文献载体,因此不属于电子文献范畴。

数字文献或网络文献只是电子文献的不同表现形式或不同称谓而已。由于电子文献以二进制数字代码形式记录于磁带、磁盘、光盘等载体上的,故人们又称之为数字文献;由于电子文献可以在通信网络上传输或在线阅读,故人们又称之为网络文献。虽然数字文献或网络文献与电子文献在概念的内涵与外延上有些差别,但不可否认的是,数字文献或网络文献都属于电子文献范畴。

1.1.2 电子文献的特征

电子文献既然是以电子化形式将文字、图像、声音等存储在光、磁等存储介质(如 CD-ROM、磁盘)上,并能通过计算机、通信设备再现出来的信息资源,那么,除了上述所说的“两个基本要素”或“两个最基本的特性”之外,相比传统文献而言,它还表现出如下几个明显特征。

1. 文献资源的数字化

电子文献既然是现代科技发展的产物,能够将各类传统介质的文献进行压缩并转化为数字信息,那么它的存在就带有明显的技术特性,需要一定的技术支撑。电子文献的数字化特性主要表现在:一是将知识和信息转换成电子形式的存在;二是要实现对于这些电子形式文献的处理、传递和使用;三是必须依赖于一定的具有高技术含量的“电子设备”,包括计算机及其相关外部设备和软件、磁盘、光盘等存储介质,以及计算机网络等必不可少的物质基础。也就是说,这种高技术特性既表现在“硬”的方面,又表现在“软”的方面。但是,其总体发展趋势是朝向友好型、人性化及多元化的展现方式。

2. 文献实体的虚拟化

在网络环境下,电子文献的信息实体可以方便地转化为数字形式在全球范围内传输,任何拥有计算机终端的地方都可以利用电子文献。电子文献的虚拟化使它突破了传统文献的信息记录规模,使文献具有超常的扩展性,使超大规模