

●刘延章

关于网络信息分类组织研究中的几个问题

摘要 《中图法》不适于组织网络信息,应建立全国统一的网络信息分类法。编制网络信息分类法应摒弃以科学分类为依据的原则,必须把实用性和易用性放在首位。只有自然语言同人工语言相结合,才能编制出高质量的网络信息分类法。参考文献 35。

关键词 网络信息分类 搜索引擎 文献分类法 自然语言

分类号 G254.11

ABSTRACT The author thinks that the *Chinese Library Classification* is not suitable for the organization of network information, and we should have a classification system for the network information in China. We should put our emphasis on the practicality and usability when compiling such a classification. The author also thinks that only by the integration of natural and artificial languages, can we draft a high-quality classification system of network information. 35 refs.

KEY WORDS Network information classification. Search engine. Document classification. Natural language.

CLASS NUMBER G254.11

被冷落多年的文献分类学如今又热起来了。它把网络信息组织作为自己的研究重点,成为图书情报界研究的一个热区。据笔者不完全统计,从1998~2002年这5年间,涉及这一领域的研究性文章有162篇。经过近几年的研究和交流,同仁们在许多方面达成了共识,但在一些重大问题上仍存在着不同的意见和看法。现就后一种情况及管见陈述于后,以引起同行的关注和讨论。

1 是改造《中图法》还是另起炉灶

笔者曾撰文指出,当前中文搜索引擎的分类体系存在着归类不够科学,序类缺乏逻辑性,以及划分不够严密,命名缺乏准确性等问题^[1]。许多学者纷纷发表文章,提出解决问题的方法和途径。有的学者认为,提高中文搜索引擎分类体系质量的途径,首要的是使用传统的分类法,也就是说根据网络信息资源的特点,对传统的分类法加以适当的调整和改进,应用到中文搜索引擎的分类目录中去^[2]。那么,使用哪一种传统的分类法?又如何去改进呢?有学者认为,以《中图法》为基础,编制出我国第一部网络信息分类法——《中图法 网络信息分类表》,才是最佳的选择^[3]。这种观点实际上是把网络信息分类表作为《中图法》的一个版本,属于改造《中图法》之列。有的学者指出,从我国的具体情况来看,对《中图法》

不断进行改造、完善,使之满足新的标引和检索需求,是一条十分现实的路子,并提出为适应计算机检索而对《中图法》进行改造的基本思路和要点^[4]。也有些学者认为,如果像一些同仁所言,组织力量编出一部《中国网络信息分类法》,用以取代《中图法》是不实际的。分析《中图法》组织网络信息的优势与不足,研究其现代化改造之措施,尽早出第五版和网络版,恐怕是一条切实可行的路子^[5]。

但也有一些同仁持不同意见。有人指出,现有文献分类体系与网络信息资源矛盾重重,解决这些问题最好的途径是编制一部有别于文献信息分类法的《网络信息分类法》^[6]。也有人指出,网络环境下传统的文献分类方法存在着一定的不适应性,建立新的网络信息分类体系势在必行^[7]。那么,如何建立新的网络信息分类体系?新的网络信息分类体系又是什么样子呢?石晓华、王春芳撰文指出:“根据以上对传统分类法和现有网络分类法优劣长短的分析,我们应该把两者加以改造和利用,编制出一部完善的科学的网络信息分类法。”并提出了编制网络信息分类法的具体设想^[8]。张燕飞老师也撰文指出:“对于分类搜索引擎来讲,可吸收传统分类法的优异性能,结合网络信息的特点,构建一个结构简明、条理清晰、层次清楚的知识体系”^[9]。

笔者认为,相对网络信息分类体系而言,《中图法》已经患上了结构庞大、缺乏简明性,划分太细、缺

乏实用性,操作复杂、缺乏易用性,体系僵化、缺乏灵活性,单线排列、缺乏多维性等综合顽症^[10],连标引物理载体文献都“山重水复疑无路”,而面对海量的网络信息焉能“柳暗花明又一村”?如果像有的学者所言,以《中图法》为主,采用《中图法》的体系结构,同时根据网络资源的特点加以调整^[11],那么,以科学分类为基础的《中图法》,怎样调整才能成为以事物分类为基础的网络信息分类法?显然此路不通。因为二者有着本质的不同。正像张琪玉教授指出的那样:“以图书情报领域使用的传统的文献分类法模式来改造因特网大众分类法是很难行得通的(更不可能用传统的文献分类法来取代其职能)。”^[12]如果像有的学者所言,对《中图法》进行大手术,变原来的硬分类体系为软硬结合的分类体系,使其重建^[13]。那么,重建后的分类法已不再是《中图法》而是别的什么分类法了。因为后者较之前者已经发生了质的变化,是此而不是彼了。

2 是建立统一大法还是各行其事

目前一些中文搜索引擎分类体系的构建的确十分粗率,其严密性和逻辑性都很差,类名措词随意性大,列类五花八门,甚至使人无所适从,到了无法容忍的地步^[14]。那么,是什么原因造成如此的局面呢?又如何改变这种局面呢?有学者指出,导致这些缺陷的主要原因是因为设计这些网络信息分类体系的大多数是一些计算机技术人员,他们当中大多数不具备情报语言学的专业知识,不熟悉传统的文献分类法^[15]。也有的学者认为,主要原因在于这些搜索引擎绝大多数采用自编分类体系,各行其事,因此,建立统一的网络信息分类体系有着重要的现实意义^[16]。还有些学者指出,用户在查询信息时,一般需要使用多个搜索引擎。但是不同的搜索引擎使用不同的分类体系,而且即使类名相同,外延也不一样,势必要求用户熟悉各种搜索引擎的分类体系,并且造成用户理解和运用的困难。因此,建立统一的中文搜索引擎分类体系势在必行^[17]。

但许多专家对“建立统一的中文搜索引擎分类体系”有异议。有的专家认为,我们不可能设计一个统一的分类体系,要求现有的所有搜索引擎抛弃自己原有的分类体系推倒重来使用这个新体系。这对于个性化强烈的诸网站来说是不可能的,因此“统一”这条路是走不通的^[18]。也有的专家认为,由于不同层次、领域、地区的网络用户对检索存在着不同的要求,编制特色各异的分类网络系统是必要的^[19]。

笔者拙见,虽然不同的信息网站拥有不同的索引数据库,因而对标引和检索工具的要求也不尽相同,但这也正像不同性质的图书馆收藏着不同性质的文献一样,对分类法的要求也不一样,不也是照样统一使用一种分类法(如《中图法》)吗?况且要想改变目前一些中文搜索引擎分类体系各自为政、无所适从的混乱局面,只有走“建立统一的中文搜索引擎分类体系”这条路子,别无选择。那么,如果建立“统一大法”之后各个网站仍然我行我素怎么办?这就要学习《中图法》宣传、推广的经验了,当然最好作为国家标准加以推行,就像《中图法》目前作为国家试行标准那样。

3 是强调科学性还是强调实用性

我们在前面曾多次提到目前中文搜索引擎分类体系存在着诸多问题,特别是科学性差的问题尤为突出,于是乎众多学者便不约而同地都把科学性作为编制网络信息分类法的首要原则而加以强调。有学者认为,制定网络信息分类法的原则之一是以科学分类为依据^[20]。也有的认为,与传统分类法一样,网络信息分类法也必须具备科学性原则,分类体系仍必须以科学的知识分类为基础^[21]。还有许多学者也都主张把科学性作为构建网络信息分类法的首要原则,必须加以遵循^[22~24]。

但也有一些学者有不同见解。王忠红指出,在网络环境下,用户获得信息的方便性和易用性越来越受到重视^[25]。因此,史学斌认为,编制网络信息分类法首先应遵循的也是对用户来说的实用性原则^[26]。我国著名的文献分类学家陈树年先生一针见血地指出:“分类法是一种工具,其实用性是第一位的。”^[27]

那么,究竟科学性是第一位,还是实用性是第一位的呢?大家应该记得,20世纪五、六十年代,关于我国图书分类法编制中“三性”(即思想性、科学性、实用性)原则的讨论,可以说是连篇累牍。讨论的结果是:《中图法》严格按照“三性”原则及其次序来构建自己的分类体系,从而导致《中图法》遗患无穷,难以根治。单就科学性而言,由于《中图法》强调“分类体系要符合科学性的原则,以科学分类为基础”(见:《中图法》第一版编制说明),因此,用户想要查询个别语种的话必须按照语系→语族→语支→语种的顺序来翻检,倘若不知道要查的语种属于哪个语系、语族、语支的话,就无从下手。除了语言学家又谁能知道吐鲁番语和库车语属于《中图法》的“H₇ 印欧语系”→“H₇₉₅ 吐火罗语族”

呢?因此,王忠红、张涵曾指出,传统分类法在用户定位方面有两个误区:一个是强调类表的科学性和专业性,忽略类表的易用性;另一个是把中介用户—分类标引人员作为最终用户,忽略检索者的需求与利益^[28]。高薇薇、杨丽认为,作为一种用概念揭示主题内容的检索工具,如果不能发挥其应有的功效,那么,这种工具就要被改造,直至被淘汰。同时又认为,《中图法》是以科学性为标准的,网络信息分类是以实用性为标准的,这是前者不能作为后者最佳选择的原因所在^[29]。真可谓一语中的。

4 是使用自然语言还是使用人工语言

我们知道,传统的文献分类法使用的是典型的人工语言^[30],而网络信息分类法使用的主要是自然语言^[31],因此有人就认为,在网络环境下采用自然语言检索是发展的大趋势^[32]。也有人认为,人工语言严格的规范性使用户丧失了检索中的自主性和能动性,背离了以人为本的原则,因此,自然语言是检索语言发展的必然^[33]。

对目前许多研究者根据人工语言不及自然语言以及计算机技术可以解读一切问题的信念,因而断定“人工语言在网络环境下已不再是有发展前途,最终将被自然语言取而代之,因而不具有价值”这一偏见,张琪玉教授指出:“网络信息检索不能唯一地使用自然语言,也必须使用人工语言。”他又精辟地指出:“情报检索用语发展的大趋势,是情报检索语言的自然人语言化和自然语言的情报检索语言化,是两者的初级结合到完全融合的过程。”^[34]陈晶也指出:“改进受控语言的易用化是网络环境下的大势所趋,而对自然语言进行必要的控制也是网络环境下势在必行的,两者的相互渗透、有机结合是情报检索语言发展的大趋势。”^[35]

据笔者考察,目前的中文搜索引擎分类体系并没有、也不可能单纯使用自然语言,因为有时候自然语言与人工语言是一致的。问题在于使用自然语言使人费解、不能表达涵义时,而没有改用人工语言。因此,笔者认为,在编制网络信息分类法时,尽可能使用生动活泼的自然语言,少用死板僵化的人工语言,但在使用自然语言不能明确表达类目的内涵和外延时,才使用人工语言。另外,还可设置 Web 页,建立帮助系统,列出自然语言与人工语言对照表,以备网民遇到自然语言不能说明问题时,求助于人工

语言。

5 几点简单的结论

《中图法》不适合用于组织网络信息,即使是对其进行现代化改造也不可能取代目前各种自编的网络信息分类体系。

建立全国统一的中文通用网络信息分类法是当务之急,是解决目前许多中文搜索引擎分类体系存在的问题,从而提高检索效果的最好措施和唯一途径。

我们不能再走以科学分类为依据编制传统文献分类法的老路,编制网络信息分类法必须把实用性和易用性放在第一位。

单纯使用自然语言或单纯使用人工语言都会降低网络用户的满意度,只有将二者有机地结合在一起,才能编制出高质量的网络信息分类法。

参考文献

- 1 刘延章,谢琳惠. 中文搜索引擎分类体系的特点、问题及其改进. 郑州大学学报(哲学社会科学版),2002(6)
- 2 焦薇玲. 论提高中文搜索引擎质量的途径. 情报科学,2002(1)
- 3 贺定安. 关于编制《中图法·网络信息分类表》的构想. 国家图书馆学刊,2002(3)
- 4,27 陈树年. 面向 21 世纪《中图法》及其标引技术改造的研究. 见:戴维民等主编. 面向 21 世纪的情报语言学. 北京:北京图书馆出版社,2000:148~164
- 5,10 崔慕岳,刘延章,张中秋. 《中图法》组织网络信息的可行性、不适应性及其现代化改造. 郑州大学学报(哲学社会科学版),2001(6)
- 6,24 邓香莲. 文献信息分类与网络信息分类之比较研究:兼谈构建网络信息分类体系. 情报资料工作,2002(6)
- 7,26 史学斌. 网络信息分类体系. 图书馆,2002(2)
- 8 石晓华,王春芳. 网络信息分类与传统分类法的优劣分析—兼谈构建《中国网络信息分类法》. 图书馆理论与实践,2001(1)
- 9 张燕飞. 搜索引擎中分类方法的应用及其局限性. 图书情报知识,2002(3)
- 11 敬卿,吴静. 网络分类目录规范化初探. 图书馆,2002(1)
- 12,14 张琪玉. 因特网大众分类法的本质属性. 图书馆杂志,2002(11)
- 13 蔡厚勇. 论图书馆数字化过程中的信息分类体系重建. 大学图书馆情报学刊,2001(3)

(下转第 30 页)

器进行系统性的调节和控制,后果将难以想象。正因如此,笔者坚持认为,正如信息技术应用是人类社会发展的大势所趋,信息产业和信息经济是国民经济发展不可逆转的潮流和方向一样,将国家信息法列为法律体系中独立的部门法是社会发展的信息化大趋势使然,将不以人的意志为转移。更何况,从现行的信息立法实践来看,虽然已有的信息法律规范比较散乱,但所调整的对象性质是相同的,即信息领域里的各种社会关系。这些发生于信息领域的各种社会关系不仅量大、面广,而且常常纵横交错、矛盾重重。国家信息法的部门法地位正是在解决这些矛盾的过程中逐渐形成并不断提升的。

参考文献

1 <http://www.nanyang.gov.cn/xxh.htm>

- 2,7 [俄]B·A·科佩洛夫.论信息法体系.国外社会科学,2000(5)
- 3 黄瑞华等.信息立法初探.中国信息导报,1996(3)
- 4 贾文中等.试论信息法的体系.情报理论与实践,1997(1)
- 5 朱庆华等主编.信息法教程.北京:高等教育出版社,2001
- 6 吴宏亮等.论我国信息立法的几个基本问题.中国图书馆学报,1999(2)
- 8 马克思.马克思恩格斯选集(第四卷).北京:人民出版社,1973:162
- 9 马克思.马克思恩格斯全集(第三卷).北京:人民出版社,1960:378

查先进 武汉大学信息资源研究中心、信息管理学院副教授、博士.通讯地址:武汉市.邮编 430072.

(来稿时间:2002-08-19)

(上接第 18 页)

- 15 罗雪英,陈代春.网络信息分类的现状与展望.现代情报,2002(11)
- 16,21 吴丹.网络信息分类体系设计.图书情报知识,2002(5)
- 17 谭宇红,胡德华,柳晓春.中文搜索引擎分类体系研究.情报科学,2001(6)
- 18 侯汉清,薛春香.中文搜索引擎分类体系兼容互换工具的设计.中国图书馆学报,2003(1)
- 19 马张华.文献分类法在网络资源组织中的应用.见:戴维民等主编.面向 21 世纪的情报语言学.北京:北京图书馆出版社,2000:165~172
- 20 曹东,韩全惜,庄军.运用情报检索语言理论与方法完善搜索引擎的功能.情报理论与实践,2000(1)
- 22 邱燕燕.网络信息的分类及对策.图书馆论坛,2001(2)
- 23 傅欣,李伟.网络信息分类组织发展趋势探析.图书情报工作,2002(3)
- 25 王忠红.网络环境下的传统分类法.图书情报工作,1999(2)
- 28 王忠红,张涵.网络环境下分类法的变革与发展.见:戴维民等主编.面向 21 世纪的情报语言学.北京:北京图书馆出版社,2000:173~180

- 29 高薇薇,杨丽.网络信息分类体系与网站分类目录初探.图书馆建设,2002(3)
- 30 陈树年.面向用户、面向新的知识载体、面向新的技术环境,发展我国 21 世纪的情报检索语言.见:戴维民等主编.面向 21 世纪的情报语言学.北京:北京图书馆出版社,2000:9~12
- 31 张琪玉.积极为自然语言与情报检索语言的结合创造条件——建议大量编制自然语言词表.见:戴维民等主编.面向 21 世纪的情报语言学.北京:北京图书馆出版社,2000:199~206
- 32 倪莉.网络信息的自然语言检索.图书馆理论与实践,2002(4)
- 33 栗慧.以人为本的必然:人工检索语言向自然语言的转变.图书馆,2000(2)
- 34 张琪玉.网络信息检索语言的发展趋势(RP)(16KB 文件).见:张琪玉论文集(光盘)
- 35 陈晶.论网络环境下情报检索语言的发展.情报杂志,2002(6)

刘延章 郑州大学信息管理系教授.通讯地址:郑州市大学路 75 号.邮编 450052.

(来稿时间:2003-04-03)