

网络信息分类浅议

黄 建 年

(江苏财经高等专科学校图书馆, 镇江 212413)

摘要 网上信息种类繁多,科学的分类相当重要。文章对网络文献信息与传统资料的特点进行了比较,并根据目前流行的 YAHOO、SOHOO 两大搜索引擎采用的分类方法,指出目前网络信息在分类中存在的问题,认为编制网络信息分类法刻不容缓。

关键词 网络资源管理 网络信息分类 分类法

On Internet Information Classification

Huang Jiannian

(Library of Jiangsu College of Finance and Economics, Zhenjiang 212413)

Abstract This article describes the features of Internet information and analyses the disadvantages of current classifications on Internet information, addresses how to design a perfect classification on Internet information.

Key words classification, Internet resources management, Internet information classification.

1 概 述

近年来,中文网络站点发展迅速。据不完全统计,中文网络站点的总页数已经超过 70 万个。如此庞大的信息宝库,任何一个网民都无法在其中“漫游”,于是对于中文信息搜索引擎的需求日益高涨起来。为了适应这一需要,国内外网络信息服务商纷纷推出了自己的中文网络搜索引擎。一时间,雅虎、搜狐、悠游、天网、网易、搜索客、华好网景等搜索引擎成了网民们上网必用的工具,甚至在网民之间流传着“出门靠地图、上网靠搜狐”这样的口头禅,这不能不说是网民们的幸事。但是,我们究竟需要什么样的搜索引擎?现有的搜索引擎基本上采用了分类查询的方法,它们所采用的分类法是否已经完善?如果已经非常完善,为什么图书馆界长期以来引以为豪的分类理论和方法,不能对此有所贡献?如果还没有完善,那么图书馆界究竟能不能对网络信息的分类有所贡献?现有的搜索引擎没有一个是图书馆界设计出来的,应该引起我们的注意。图书馆界能否编制一部用于网络信息分类的“网络信息分类法”呢?要想了解这一切,我们有必要对网络信息与传统信息的异同以及现有搜索引擎的优缺点做一次较为深入的探讨和研究。

收稿日期:1998 年 12 月 2 日

作者简介:黄建年,男,1966 年生,北京大学图书馆系情报学系毕业,馆员,主要研究图书馆自动化及数据库建设。

2 与传统资料的异同

2.1 活动性

一般说来,网络信息变化较快。传统资料中更新更快的莫过于报纸了,但是报纸一经出版其信息即无法变动。而网络信息则不同,它不仅可以快速出版,而且可以根据用户的意见及时作出变更。

2.2 无序性

网络中既包括有序化的信息,也包括无序化的信息。比如在 USENET 中就有很多信息未经任何整理,无序程度相当高。但是,需要注意的是,也正是这些无序性信息,有时却代表了最新的科研动态和科技发展水平。对这种信息的利用,应该引起我们足够的重视。

2.3 扩充性

传统资料一经出版,就有了固定的样式,无法加以扩展和扩充。网络资料则不同,可以根据实际需要,随时加以扩充。现在网络中有些刊物,没有固定的版面,可以根据收稿量的多少及时加以调整。

2.4 可选择性

网络刊物可以根据用户的需要设立不同的版式和收费标准,以方便用户订购和使用。而传统资料则很难做到这一点。

2.5 暂时性

传统资料一旦出版,如无特殊情况一般可以长久保存下来,而网络信息则不同,有不少信息属于“过眼云烟”。比如股市行情、会议信息等等,变化非常迅速,经过一段时间的网络传播后,一般不会保存下来。

2.6 关联程度强

传统资料的关联能力较差,主要通过参考文献、引用文献等方式来指导读者阅读和核对所引用的资料,如果用户找不到所引用的原始资料,那么就无法对某一资料进行详细地分析和研究。而网络信息则不同,它可以利用超文本技术,通过超级链接手段直接指导用户去查阅所引用的原始文献。

3 传统图书资料分类法的局限

网络信息有上述这些特点,传统的分类法能否适应这种需要呢?笔者认为,传统分类法在这方面的功能有所欠缺,很难胜任,具体表现在以下几个方面:

3.1 更新慢

传统分类法更新比较慢,这是众所周知的事实。就以《中图法》为例,到目前为止,才出版

了第四版,因而对新资料分类的不适应性也就显而易见。关于《中图法》分类中存在的问题屡有论文见诸报刊之中,不需笔者多论。

3.2 变更难度大

每一次分类法的改版,一方面给读者和文献分类人员带来了方便,另一方面则又给图书馆带来了工作量巨大的文献分类、目录的改编工作。为此,一些图书馆采用了“分段式”的处理方法。这种处理方法给读者使用图书馆目录带来了极大的不便。

3.3 表达关系的能力有限

传统文献分类法在表达主题、类目之间关系方面主要采用参见、组配等方式,比较适合传统文献主题相对单一、学科分野相对清晰的特点。但是随着科技的发展,学科分野比较模糊,边缘学科、交叉学科增多,现代文献中属于多主题、边缘学科的比比皆是,仍然采用传统的参见、组配方式显然已经不能适应现代文献分类的需要了,更遑论能适应变化万千的网络文献信息的分类了。

4 YAHOO !、SOHOO 所用的分类法

为了方便用户查找网络信息的需要,目前的网络搜索引擎大多有自己的分类法。这些分类法怎样?是否已经满足网络信息检索的需要了呢?作一个简单的介绍很有必要。

4.1 中文 YAHOO ! 所用分类法概述

作为世界著名的搜索引擎,YAHOO ! 相当有自己的特色,在分类方面也是如此。中文 YAHOO ! 的分类索引共分为休闲与运动、区域、社会科学、教育、社会与文化、参考资料、政府与政治、商业与经济、科学、新闻与媒体、娱乐、电脑与因特网、健康与医药、艺术与人文等 14 大类,基本上包括了人类生活的各个方面。

4.2 SOHOO 所用分类法概述

作为全中文的搜索引擎,SOHOO 的发展非常迅速,目前已经“达到 7 万以上的日访问人数和 40 万的页读数”,成为“互联网上的第一中文网站”。SOHOO 为何能取得如此辉煌成就?其秘密在何处?用 SOHOO 自己的话说,就是采用了“最先进的人工分类技术”,并且“符合中文语言文化习惯”。可见,分类技术在此网站的发展中所起的重要作用了。

SOHOO 将所有网络信息资源共分为娱乐休闲、工商经济、计算机与互联网、新闻与媒体、科学与技术、文学、艺术、卫生与健身、体育与健身、生活服务、社会科学、旅游、教育、社会与文化、地区、综合参考、科学与宗教、政治与法律等 18 个部类,也囊括了人类生活的各个方面。

4.3 YAHOO 法和 SOHOO 法的贡献

4.3.1 综合性信息网络分类体系

无论是中文 YAHOO !,还是 SOHOO,其分类索引都涉及了社会生活的各个方面,各种类型的网民都可以利用它们查找到自己所需的资料和信息。目前,许多网站的信息分类基本上采用了上述二者的分类体系。可以这样说,YAHOO !、SOHOO 采用的分类法实际上已经成为

目前类分网格信息的新的分类法,我们不妨将之称为“YAHOO法”和“SOHOO法”。

4.3.2 面向中文信息的网络分类体系

无论是中文 YAHOO! 还是 SOHOO,都以互联网中的中文信息站点为主要数据来源。研究二者的分类体系,并加以完善,将使我们产生一部真正意义上的、完善的中文信息网络分类标准。

5 网络信息分类存在的问题

现有的搜索引擎在分类方面取得了令人瞩目的成就,值得图书馆界、情报界很好地加以研究。但是,无论是 YAHOO法,还是 SOHOO法,其编制者均非长期从事科学分类工作的图书信息方面的专家,自然所编制的分类法也就或多或少地带有这样那样的问题。对这些问题的分析,必将有助于网络信息分类法的完善。

5.1 体系不统一

即使在同一部分类法中,体系也不尽统一。比如,中文 YAHOO! 法的基本大类中,“电脑与因特网”却与“科学”并列,SOHOO法的基本大类中,“计算机与互联网”中“科学与技术”并列,这些作法,很显然不符合科学分类的有关规定。

5.2 类名不规范

在网络信息分类法中,类名不规范已经是众人皆知的事实。在上节的事例中,就有两处不统一:其一,YAHOO! 称“电脑”,SOHOO 称“计算机”;其二,YAHOO! 称“因特网”,SOHOO 称“互联网”。作为主题,“电脑”、“互联网”均可,但是作为类名,显然欠妥。关于“INTERNET”的命名问题,国家有关部门早就下发了有关文件,统一称为“因特网”,SOHOO 仍然称为“互联网”,显然不够规范。

5.3 分类缺少提示

用户不能直接找到所需的类名,必须一级一级地向下翻阅,有时为了查找一个类名,需要读取十多个页面,费时费力,使用起来十分不便。最好能有一个电子版本的分类法,供用户下载、熟悉、研究之用,以便上网时能够快速找到所需的类名。

5.4 无分类代码

传统分类体系中,每个类名都有一个对应的代码,用户在检索时如果知道分类代码,查找起来十分方便。但是,现有的网络信息分类都没有分类代码。这种做法,对于经常使用某一类信息的用户十分不便。如果能够加上分类代码,将给用户带来莫大的好处。

5.5 层次不尽合理

有些类目只有两三层,而有些类目则多达十多层。作为分类法,其类目不是不可以达到十多层,关键在于所有类目的设置是否合理。笔者以为,各大类下的层次应该基本上相近。也就是说,现有的网络信息分类体系应该进行一些适当的调整。

6 网络信息分类的基本原则

了解了现有网络搜索引擎存在的问题,我们就能够有的放矢地对现有的网络分类法进行改进,并且充分利用传统分类法的优点,编制一部更加科学的、完善的网络信息分类法来。笔者以为,网络信息分类法至少应该遵循下列原则:

6.1 直接性原则

分类的层次不宜太多,不应该出现转换十多次还找不到所需信息的情况。一般来说,在6层以下的类目比较符合人们的习惯。笔者在实践中发现,如果类目的层次超过6层,用户就会改用其他检索途径,比如主题词等。编制网络信息分类法时必须注意这一问题。

6.2 实用性原则

网络信息分类的类目要少而精,要有用,不应该仅仅为了美观因素而编制一些无用的类目。关于这一原则,在传统分类法中有比较详细的解释和研究,此处不再赘述。

6.3 自然性原则

网络是面向大众的,所以,采用的类名基本上应是自然语言,以满足广大普通用户的信息需求。换句话说,网络信息分类的类名不应学术性、专业性太强。YAHOO!法、SOHOO法中都很少用“×××理论”、“×××学”这样的类目,就是明证。比如《中图法》中与经济有关的类目被称为“经济学”、“财政学”、“投资学”,而在YAHOO!法中则被称为“商业与经济”、“金融与投资”;在SOHOO法中,则被称为“工商经济”、“金融与投资”、“保险”等,无一与“学”有关,可见其立类原则了。我们要编制一部网络信息分类法也必须充分研究各大搜索引擎所用的类名。

6.4 快捷原则

每一类名应有对应的一个快捷键或一个类号,以便于熟练用户快速查找。传统分类法在这方面有相当杰出的成就,编制网络信息分类法毫无疑问地应该加以研究和借鉴。

7 结 语

关于网络信息的分类,目前研究还不多,尚没有出现一部严格意义上的分类法。笔者真心期望拙文能够起到抛砖引玉的作用,引起诸位专家学者对网络信息分类问题的重视,并衷心祝愿通过图书信息界的共同努力,能够早日编制出一部真正的网络信息分类法来!

参 考 文 献

- 1 <http://www.yahoo.com>
- 2 <http://gbchinese.yahoo.com>
- 3 <http://www.sohoo.com>
- 4 中国图书馆学分类法(第三版).北京:书目文献出版社,1991
- 5 中科院文献情报中心.文献分类.北京:中科院文献情报中心,1989
- 6 北京大学图书馆学情报学系.图书分类.北京:书目文献出版社,1990