

潘燕桃

地方报纸信息数据库建设存在的问题及发展设想

ABSTRACT There has been a rapid development in the creation of databases of local newspapers. Some of them have already been put on the Internet. However, there are still some problems such as lack of cooperation, low quality, lack of local characteristics and inefficient search methods. The author proposes some recommendations to solve these problems.
1 tab 4 refs

KEY WORDS Local newspapers Database development

CLASS NUMBER G356

在我国报纸信息数据库建设中,地方报纸信息数据库发展较快。例如广东省的《南方日报》《广州日报》《东莞日报》等不但建立了报纸信息数据库,而且还继《人民日报》之后进入了因特网,向全球推出了电子版,使国外的华人同胞能够及时地了解广东在改革开放中的新风貌、新变化。

但是,由于政策与机制、技术手段与设备、社会信息意识、人员素质等多方面的原因,我国地方报纸信息数据库的建设尚存在许多问题。本文通过对上述地方报纸信息数据库的现状与存在问题的分析研究,提出一些设想和改进措施,以期对今后地方报纸信息数据库、电视台、电台新闻数据库以及其他数据库的建库工作实践提供一点参考和借鉴。

1 地方报纸信息数据库建设存在的问题

1.1 宏观调控不力,数据库建设重复现象严重

由于国家、地区间的宏观调控不力,主观上缺乏统一规划的思想意识,技术上缺乏网络环境的支持,属于事业单位的各报社缺少需求分析和用户界定,也很少注重数据库的利用率问题,致使各地方报社自行投资建库,

造成数据库规模小,缺乏竞争机制,建库方式单一,在选题上重复与空缺并存,造成人力、物力、财力的极大浪费。例如《南方日报》《广州日报》和《羊城晚报》,各自都建有报纸信息数据库。由于同属于广东地区的地方报纸,相当一部分内容是重复的。像“中国新闻”、“国际新闻”、“广州新闻”、“体育新闻”、“天气预报”、“金融证券信息”等报道,因为消息来源是一致的,所以几乎每天的内容都大同小异。

目前,从总体上看,地方报纸信息数据库开发重复现象严重;而从个别来看,由于地方报纸信息的地方性、复杂性与时效性和技术支持不够,以及人员专业素质较低等因素的影响,普遍存在现有的报纸信息资源开发不足的问题。

1.2 检索效果未如人意

一个实用可行的报纸信息检索系统应是外部特征检索、主题检索、分类检索、全文检索的综合应用^[1]。目前已经建立的地方报纸信息数据库大多都只有外部特征检索和全文检索两种检索途径(见表1)。

一般的用户都能快速简便地从作者、标题、日期、版次等外部特征入手查找资料,但前提是用户已事先知道要查找的资料所在的

表 1 广东部分地区报纸信息数据库可检字段一览

单位	外部特征字段	内容特征字段
广州日报	(起始)时间、时间版次、作者、标题	关键词
南方日报	(1993—1996)日期、作者、标题、栏目刊头、头版头条、版名、版次、体裁、版序、类目 (1997—)日期、标题、版次、版名	类目名称 关键词
羊城晚报	(1991—1993)(起始)时间、时间版次、作者、标题 (1994—)标题、作者、时间、专栏、版次、注释(专题报道名称等)	专栏名称 关键词
深圳特区报	标题、作者、时间、版次、地区、来源	关键词

日期及其他信息,否则,用户要想扩大或缩小检索范围,查全查准资料,并非易事。数据库缺乏科学的辅助检索工具,影响了它的使用效果。

全文检索与其他检索途径相比,虽然具有不需人工标引加工、便于查找文章细节、检索结果直接可靠、检索详尽彻底、使用简便易学等优点,但在不进行任何控制的情况下,文章中的每一个字、词均可成为检索点,虽在输入方面大大节省了人力和存贮空间,却加重了检索方面的负担,使每次检索命令中的文献量偏多,筛选量过大,用户检索所需时间长,直接影响了查准率,造成了检索噪音。再加上“假义词”、同义词大量存在,更容易造成漏检错检,难以保证查全率和查准率。

1.3 人员素质及其知识结构影响数据库发展进程

由于多方面原因,我国新闻资料从业人员整体素质不太高,知识结构不合理。据调查统计,具有高级职称的新闻从业人员仅占全部人数的 5%^[2],信息学情报学专业毕业的工作人员更是凤毛麟角。近年来,一些沿海经济发达地区,尤其是广东地区的新闻资料从业人员的大专以上学历的比重有所增加,但大部分属于后继学历,而且学科专业不对口,从整体而言,同样存在素质不高、知识结构不合理的问题:负责开发数据库的计算机技术人员一般都不甚了解数据库所属专业,而数据库所属专业人员反过来也不懂计算机技术;负责开发报纸信息数据库的计算机技术人员,一般都不熟悉报纸信息的分类、主题标

引等业务工作,同样,从事报纸信息业务的工作人员通常都不懂得计算机程序编写及系统设计等知识,或是只懂得一些计算机的基本操作。因此,常常囿于双方的沟通和知识结构差异的问题,都会不可避免地、或多或少地影响数据库的发展进程。

1.4 商品化程度尚低,不能向社会开放

目前我国已建的地方报纸信息数据库大多只供本单位内部使用,一般不能向社会开放,能提供商业化服务的更是屈指可数。其原因是:

(1) 地方报社的行政体制束缚了信息服务的社会化与商品化进程。

(2) 信息市场发育水平低,信息的价值与价格关系没有理顺,地方报纸信息服务的社会化与商品化缺乏大环境。

(3) 社会信息化意识不强。一方面建库者对数据信息的社会性认识不足,只想独霸,不愿共享;另一方面是社会公众缺乏利用现代化手段获取信息的意识,习惯于从行政渠道下达的政策文件中获取信息,形成了信息需求上的惰性,即使是横向联系也要通过上级管理机构进行协调;加上对地方报纸信息数据库了解不多,或是由于网络通信等问题而放弃使用数据库。

(4) 缺乏必要的资金投入与人力投入。数据库是属于技术知识密集、资金密集和劳动密集型的事,没有足够的资金、人力及技术力量的投入,难以快速启动建成并维持长期正常运转和提供服务。目前多数地方报社资金仍然比较困难,人手短缺,信息数据库的工作人员

大部分都是身兼数职, 常常忙于内部服务, 很难再向社会开放或开展商业化服务。

1.5 未能突出地方特色

大部分的地方报纸信息数据库, 都只是把报纸上所刊载的内容收录入库, 而没有进行深层次开发, 没有把地方报纸信息数据库的地方特色发掘出来。目前我国地区经济发展水平不一致, 尤其是有一些特定的地区, 一些经济发展互动地区, 这些不同的地区具有不同的地域文化(包括民俗和历史)、不同的人才群落、不同的文化研究机构、不同的信息资源、不同特色的地方产品等等。所有这些都会在地方报纸上有所反映, 如果在建设地方报纸信息数据库时, 能把这些地方特色发掘出来, 就能更好地为新闻工作者乃至全社会提供服务。

2 地方报纸信息数据库发展的设想

在因特网席卷全球, 我国“十大‘金’字号国民经济信息化工程”建设实施的形势下, 笔者针对目前我国地方报纸信息数据库建设的现状及存在问题, 提出如下的发展对策与设想:

2.1 采取地区联合建库方式

联合或合作建库应是信息网络时代的首选和长期的建库方式^[3]。地方报纸信息数据库更应如此。

由于地方报纸的地方性与部分新闻来源的一致性, 同一地方的地方报纸的许多内容都大致相同。如果随着地方报纸信息数据库服务的社会化, 社会各界对地方报纸信息的需求也会越来越大, 人们期望在一个数据库中就能查阅本地区其他报社的报纸信息。在这方面“教育文献数据库”就是一个联合建库的例子。“教育文献数据库”是在广东省高等教育厅教学处领导下, 成立建库领导小组, 联合中山大学图书馆、佛山科技学院图书馆、广州师范学院图书馆等 11 所高校及师范院校图书馆共同参与建设, 分散建库, 数据共享。该库是“广东省高校图书馆网络系统工程”的

一个重要组成部分, 收录全国重点教育刊物上发表的文献, 重点收录广东和港澳台出版的教育刊物(包括有关教育方面的地方报刊和年鉴), 为各级行政部门、教育科研机构和各级各类学校的教育教学与科学研究提供教育信息。该库的建库领导小组定期召开成员馆主管馆长及建库工作人员工作会议, 协调建库工作; 举办培训班, 培训各成员馆负责标引和录入人员, 以保证数据库的质量。从 1995 年底开始建库至今, 共录入数据 15000 余条, 数据形式包括索引、文摘和部分全文, 定期出版书本式索引文摘, 提供教育信息。它不但得到了国家教委和广东省高教厅的大力支持和充分肯定, 而且还受到了国内外专家同行的好评。

“教育文献数据库”的经验表明: 联合建库是今后我国数据库建设应该选择的一种建库方式。因此, 地方报纸信息数据库建设应该采取以下模式: 在国家有关部门的宏观调控下, 在各省市委宣传部的直接领导下, 联合本省、本地区的地方报社共同建设地方报纸信息数据库, 分工建库, 数据共享, 同时注意突出各报的特色; 并通过信息网络与本省市的其他专业报社、其他各省市地方报社和专业报社联通, 组成全国报纸信息网络, 进入覆盖全国的信息国道, 充分开发和利用地方报纸的信息资源。

2.2 建设若干个实用的子库

地方报纸信息, 是一种具有浓郁地方特色、内容庞大、时效性强、类型繁杂、各种新名词、新术语不断涌现等特点的综合性信息。为了方便建库和用户检索, 根据实用性原则, 地方报纸信息数据库可建成若干个子库, 如人物数据库、会议数据库、经济新闻数据库、法律法规信息数据库、体育信息数据库、广告信息数据库、文娱信息数据库等。

2.3 突出地方特色

地方报纸信息是全面反映某一地区政治、经济、科教文化的知识宝库和信息源。随

着改革开放的深入发展,它对地方经济社会发展的咨询决策作用、物质文明建设和精神文明建设的推动作用日益增强。发掘地方特色应该是地方报纸信息数据库开发工作的重点。如广东毗邻港澳,与港澳形成多形式、多层次、多领域、多元化的经济合作关系,广东在改革开放中市场经济体制占有明显的优势;同时又具有商贸优势,一向是华南的商贸中心;并拥有大批来自全国各地的“南下人才”。所有这些得天独厚的优势,都是广东特色。应当把这方面的报纸信息资源开发出来。《南方日报》社资料室从 1996 年开始在这方面做出了大胆的尝试。

2.4 扩大收录范围,满足多层次、综合性的社会需求

作为地方报纸的专门信息数据库系统,新闻无论大小最好全部收录,要为查阅资料提供可靠的依据,力求做到只要是报上刊登过的,数据库都应能提供线索,找到具体内容,所以应该保证一定的完整性。

随着社会信息化程度的不断提高,地方报纸信息数据库的开发利用势必会越来越受到重视,新闻工作者及全社会对地方报纸信息的需求也会向综合性、多层次发展。地方报纸信息数据库必须将收录范围尽可能地扩大,才能满足需求。

2.5 建立实用的检索系统

一个实用的地方报纸信息检索系统应是外部特征检索、主题检索、分类检索、全文检索的综合应用。目前主题检索和分类检索仍比较薄弱,甚至空缺。

2.5.1 主题检索

建立主题检索系统的关键环节是主题标引工作。主题标引的一致性、准确性直接影响文献检索系统的查全率和查准率。主题标引是对事物的特定对象及其各个方面的问题进行研究和探索,它可集中同类之中的相关主题,人们可以从不同的角度对某一事物进行检索提问,解决了多主题、新概念层出不穷

等复杂资料的加工处理及检索问题,并能充分发挥计算机的先进技术。

2.5.2 分类检索

由于一直沿用的《新闻资料分类法》已经陈旧,不再适用于日新月异的报纸信息,各新闻单位有的根据本报信息的特点自编分类法;有的在《新华社分类法》的基础上根据需要进行修订,改造成本报专用的分类法。再加上部分新闻工作者不习惯或不懂得从分类途径检索,致使分类检索系统利用率不高,甚至部分原设有分类检索模块的建库单位最近都放弃了分类标引。但必须指出的是,尽管目前存在着缺乏权威的地方报纸信息分类法,也不能失去分类检索这个重要的检索入口。建议编制专门的地方报纸的分面分类法。由于现有的新闻资料分类法均不适用于地方报纸新闻资料的分类,在很大程度上影响了分类标引的进行,如果能够组织有关专家学者编制专门的地方报纸的分面分类法,将会有利于各地方报社开展分类标引的工作。

2.6 建设多媒体地方报纸信息数据库

多媒体技术集声、文、图、影像、动画信息形式于一体,不仅可以用键盘和鼠标操纵计算机,而且可用笔语言、触摸屏幕来与计算机对话,使人们与计算机的交流达到自然会话的境界。建设多媒体地方报纸信息数据库是信息时代的需要,也是新闻工作者及社会各界人士的迫切需求,如能建立多媒体的地方报纸信息数据库,必将为全社会提供更高质量、更为方便的服务。

2.7 在信息网络的环境下向社会提供服务

随着因特网的迅速全球化,以及我国“十金工程”的建设的全面展开,利用信息网络高速、优质、便捷、廉价地传输和共享信息已经从梦想成为现实,为地方报纸信息数据库向社会开放提供了非常有利的环境和条件,使信息资源共享成为了可能。

《广州日报》图书阅览中心率先向社会敞开了大门。从 1997 年元旦开(下转第 79 页)

- 8 C. King, et al, America's Best Research University? Stanford Soars in Top Ten Tournament Science Watch. 1994, 5(9)
- C. King, et al Harvard Hot in Physical Sciences as Top Ten Tournament Terminates Science Watch. 1994, 5(10)
- 10, 23, 27 E. Garfield, Citation Classics-Four years of the Haman Side of Science Current Contents 1981, (22)
- 11 C. King, et al, Scotland Harrests Bumper Crop of Highly Cited Studies on Agriculture Science Watch. 1994, 5(4)
- S. Mitton. South African Science A nem ic Serious Action Surely Advisable Science Watch. 1995, 6(3)
- 13, 19 E. Garfield, The Most-Cited Papers of All Times, SCI 1945~ 1988 Part1-4 Current Contents 1990, (7), (8), (26), (34); 1991, (21)
- 16 D. M. Boff and L. L. Hargens Are Sociologists' Publications Uncited? Citation Rates of Journal Articles, Chapters, and Books Current Contents 1993, (5)
- 17 E. Garfield, The 1000 Contemporary Scientists Most-Cited 1965~ 1978 Part1. Current Contents 1981, (41)
- 18 Don Schauder, Electronic Publishing of Professional Articles: Attitudes of Academics and Implications of the Scholarly Communication Industry. Journal of the American Society for Information Science 1994, 45(2)
- 24 P. Pichappon, Levels of Citation Relations Between Papers Journal of the American Society for Information Science 1996, 47(8)
- 25 E. Garfield, The Impact Factor. Current Contents 1994, (25)
- 26 中国科技信息研究所 中国科技论文统计与分析(年度研究报告). 1994, 1995
- 袁培国 通信地址: 南京市汉口路 22 号, 南京大学图书馆. 邮编: 210093.
- (来稿时间: 1997. 8. 21. 编发者: 李万健)

(上接第 52 页) 始, 凡订《广州日报》一年或一年以上的社会各界人士, 只需交纳 45 元的借书押金及 5 元的手续费, 均可办理该中心的借书证, 免费上机检索该报数据库, 上网浏览电子报刊和借阅图书报刊资料。迄今已有 200 多人办理了借书证, 其中以大专以上学历的人居多。可见, 不但新闻工作者需要地方报纸信息, 其他工作在各条战线的人们也有着同样的需求。尽管《广州日报》图书阅览中心的做法距离地方报纸信息数据库商品化的要求尚远, 但毕竟开了向社会提供电子版地方报纸信息服务的先河, 其经验十分值得推广和借鉴。

笔者设想, 各省市、地区地方报社联合建设本省市、本地区的报纸信息数据库, 并通过信息网络与本省市的其他专业报社, 其他各省市的报纸信息网络联通, 组成全国报纸信息网络, 进入全国的信息国道, 走向国际互联网络。在这样的信息网络的环境下, 地方报纸

信息数据库向社会开放, 提供有偿服务, 将会使地方报纸信息数据库兼得社会效益和经济效益, 走上良性循环的发展道路。

参考文献

- 1 宋明亮 我国报纸信息数据库开发的现状与对策 中国图书馆学报, 1995(1): 60~ 65
- 2 李献线 关于建设和发展我国新闻数据库的思考 中山大学硕士学位论文, 1996: 20
- 3 曹树金, 罗春荣 论信息网络时代图书馆的数据库建设和服务选择 高校文献信息学刊, 1994(3~ 4): 24~ 27
- 4 Long, A. Full-text newspaper retrieval is hard to manage: fact or fiction. Proceedings of the 9th National Online Meeting: Medford, 1988: 213~ 216

潘燕桃 中山大学信息管理系硕士毕业, 现为该系讲师。通讯地址: 广州市。邮编 510275。

(来稿时间: 1997. 8. 21. 编发者: 翟凤岐)