

大学教学视角下的专题研究型数据库建设刍议

王娟¹ 葛宜楷²

1 山东大学计算机科学与技术学院 山东 青岛 266237

2 澳大利亚国立大学计算机学院 澳大利亚 堪培拉 2601

摘要: 专题研究型数据库的基本特征是: 元数据通过研究生成; 检索路径的设置, 服务于明确的问题意识和研究目的。其功能特点包括: 与特定研究领域联系密切, 信息检索结果准确化, 解决问题相对彻底。元数据内容的精确性、完善性和数量的丰富性, 以及检索路径的多元性, 是其建设要点。

关键词: 专题研究型数据库; 历代进士登科数据库; 数据库建设

随着计算机科学技术的迅速发展, 数据库建设理论和技术不断更新, 在得到持续应用开发的同时, 也已进入大学课程, 成为了一个重要的教学知识模块。本文从大学教学的视角, 探讨专题研究型数据库的基本特征、功能特点和建设要点。

专题研究型数据库是指基于某一专题领域的研究数据而形成, 主要服务于该专题领域信息查询和学术研究的数据库。相较于原始文献型综合数据库、专题研究型数据库, 因其数据内容的专属性和检索目标设置的明确性, 以及由此带来的信息检索的针对性和便捷性, 日渐受到专业人士关注, 尤其是在文史学科领域内, 已经建成了不少具有影响力的数据库。如哈佛大学费正清中国研究中心、北京大学中国史研究中心和台北“中研院”历史语言研究所合作建设的“中国历史人物传记数据库”(CBDB), 哈佛大学费正清中国研究中心和复旦大学历史地理研究中心合作建设的“中国历史地理信息系统”(CHGIS), 台北“中研院”建设的“中华文明时空基础架构平台”(CCTS), 浙江大学龚延明教授主持编纂、古联(北京)数字传媒科技有限公司设计开发的“历代进士登科数据库”等。本文主要以“历代进士登科数据库”为例, 探讨专题研究型数据库的基本特

征、功能特点以及建设要点。

1 专题研究型数据库的基本特征

专题元数据是经过专业人员研究而形成的数据信息, 其检索路径的设置服务于明确的问题意识和研究目的。

1.1 元数据通过研究生成而非原始文献

与原始资料型数据库不同, 专题研究型数据库的元数据, 是经过专业人员研究而形成的数据信息。如“历代进士登科数据库”的元数据, 是有文献记载的历代所有登科人物及其传记。就人物传记而言, 其元数据“除著录姓名、字号、籍贯、朝代, 以及具体的登科时间(年份)、所属科目和在位帝王等信息外, 还综合进士人物的出身、登科、历官等方面的经历逐一撰写生平简介。”^[1]由此可见, 专题研究型数据库的数据形成有两个特征: 一是必须由专业学术人员研究生成; 二是生成、整理数据的周期较长, 需要耗费较多的人力和时间。如“历代进士登科数据库”由龚延明教授带领专业学术团队, 将各类可见的传世文献和出土史料进行考、辨、辑、校, 历经二十年, 最终汇集而成。

1.2 检索路径的设置服务于明确的问题意识和研究目的

专题研究型数据库建设, 往往有着明确的问题意识和研究目的。因此, 检索路径设置相对简单而明确。如“历代进士登科数据库”设置“姓名”“朝代”“科目”“籍贯”和“职官”等五项检索条件, 另外设置“姓氏导航”“帝号导航”和“籍贯导航”等三个检索路径。通过“姓名”可以检索单个登科人物的信息, 而通过“朝代”“科目”“籍贯”“职官”“姓氏”“帝号”等检索条件, 可以检索符合此条件的登科人物群体的信息。由此则可以“挖掘新事实、产生新认识”, 形成超越个人文献阅读的认知。如王瑞来评价作为“历代进士登科数据库”宋代数据来源的《宋代登科总录》^[2]时说: “有了这部《宋代登科总录》, 研究地域社会文化的兴衰和人才的消长, 研究宋代政治史、社会史中的重要人物与同年现象, 便有了基本的数据, 获得了极大的便利。”^[3]

2 专题研究型数据库的功能特点

专题研究性数据库功能具有与特定研究领域联系密切、信息检索结果准确以及解决问题相对彻底等特点。

2.1 与特定研究领域联系紧密

专题研究型数据库自设计之初,即有着明确的专题范围规定,其功能设计主要服务于该专题领域内的信息查询和学术研究。如“中国历史人物传记数据库”(CBDB)主要用于中国历史人物传记资料的查询,以及利用人际关系、空间等条件而形成人物群体关系的分析研究;“中国历史地理信息系统”(CHGIS)除了具备普通历史地图的功能外,还可为历史地理或其它方面研究提供基础数据。

2.2 信息检索结果准确化

专题研究型数据库数据具有专题性特点,且经过了专业人员的研究与整理,因此信息检索结果更加准确。以“历代进士登科数据库”为例,以“籍贯”为检索条件,检索某地域历史上的登科进士,只要输入检索词“沂州”,即可获取102条搜索结果。其中仅有两条与搜索目标不符的数据,其余100条皆为搜索目标数据,信息检索结果具有很高的准确率。

相反,利用原始文献综合数据库(如通行的“中国基本古籍库”)进行上述目标检索,无论怎样组合检索条件,都很难呈现较高准确率的检索结果,甚至无法实现对目标数据的检索。

2.3 解决问题相对彻底

集中、准确的搜索结果,使得所查询的专题数据库领域内的相关问题得到了便捷而相对彻底的解决。这是专题研究型数据库的优势所在,极大方便了该专题领域内的信息查询和相关学术研究。

3 专题研究型数据库建设要点

专题研究型数据库建设需要注意元数据收集质量、数量和建立多元化检索路径等方面问题的

内容。

3.1 元数据内容要精确、完善、丰富

元数据是数据库的内容基础,因为专题研究型数据库的元数据通过研究而生成,所以无论是从质上还是量上,对元数据都有着极高的要求。可以说,元数据内容的精确、完善程度和丰富程度,是决定专题研究型数据库建设能否成功的关键。

对于专题研究型数据库的元数据,在研究、生成过程中需要做到精益求精,不断完善。“历代进士登科数据库”在这方面提供了一些建设实例。上文所提到的《宋代登科总录》的数据来源,是傅璇琮、龚延明和祖慧编著的《宋登科记考》^[4]。我们以“范仲淹”条为例,对比《宋代登科总录》较《宋登科记考》在元数据内容上的改进。

(1)《宋登科记考》:字希文,苏州吴县人。大中祥符八年登进士第。历仕枢密副使,参知政事。终资政殿学士、知颖州。卒谥文正。

(2)《宋代登科总录》:字希文,苏州吴县人。以“朱说”名登大中祥符八年进士第。初授广德军司理参军。历仕枢密副使、参知政事。终资政殿学士、知颖州。卒谥文正。两相比较,后者添加了两处内容。关于这两处添加内容的意义,王瑞来评论说:“范仲淹到了40岁的天圣六年(1028),才归宗更名,成为日后名满天下的范仲淹,而在天圣六年以前一直使用的是母携其改嫁后继父为其起的‘朱说’之名。上引《宋登科记考》径直记作‘范仲淹’,而不提及‘朱说’之名,存在一定的问题。因为那时世上还没有‘范仲淹’这个名字。此外,从这条记载看,《宋登科记考》在记仕履上,径直为‘历仕枢密副使,参知政事。终资政殿学士、知颖州’。而《宋登科总录》则有了变化,补充了

范仲淹的初仕‘授广德军司理参军’。作为登科录,记载登科后的初授官无疑是必要的。”^[3]可见这两处添加,显著增强了“范仲淹”条元数据的精确性。通过深化研究,对元数据进行精益求精的加工,是专题研究型数据库建设前期十分必要的工作。

除对单个元数据尽可能进行质的提升、完善外,元数据总量的丰富程度也直接关系到专题研究型数据库建设的成败。因为这类数据库具有专题性,所以原则上要穷尽收集专题领域内的所有元数据。如“历代进士登科数据库”所收“资源共计10万余条,囊括隋代至清代各科举类目(包括进士、诸科、特奏名、制科等),是目前国内搜录最全面、内容最权威的登科人物数据库。”由此可以看出搜录条目(元数据)的丰富性,以及单个条目撰写的精确性,成就了“历代进士登科数据库”的高质量 and 权威性。

3.2 检索路径的多元性

在收集、完善元数据后,如何利用数据挖掘技术设置检索路径,以充分发挥数据库所收材料的查询和学术研究价值,已成为数据库建设者需要解决的首要问题。“数据挖掘起源于计算机科学,一般是指从大量的数据资料中,搜索、研究出其中隐藏的有着特殊关联性 or 规律性的研究方法,是数据库知识发现(Knowledge-Discovery in Databases)”^[5]工作中的重要环节。就专题研究型数据库而言,充分考虑数据群中不同变量的关系,尽可能多地设置检索路径,以搜索出其中隐藏着的特殊关联性 or 规律性,也是保证数据库利用效益最大化的关键环节。

4 结语

综上所述,专题研究型数据库的基本特征是元数据通过研究

浅析 2035 西部城市公共图书馆服务体系 高质量发展的策略

——以西安市公共图书馆服务体系为例

方 云

西安图书馆 陕西 西安 710018

摘 要: 本文以西安市公共图书馆服务体系为例,总结了“十三五”建设成果和差距,梳理了专家面向 2035 公共图书馆“高质量发展”的观点,分析了“十四五”社会环境和政策环境,提出了总分馆服务体系的标准化服务、全民阅读线上线下融合服务、建设适老化设施以及开展服务的易操作策略。

关键词: 西部城市;公共图书馆;高质量发展

0 引言

回顾“十三五”,展望 2035,西部城市公共图书馆服务体系如何高质量发展?笔者以西安市公共图书馆服务体系为例展开讨论。

1 “十三五”期间的进展与差距

“十三五”期间,全国公共图书馆事业作为构建公共文化服务体系的重要组成部分,取得了显著的成绩^[1],呈现出“八大亮点”^[2],在相关法律体系、服务标准、总分馆格局、全民阅读、数字化网络化建设等方面成绩斐然。但与东部相比,西部地区图书馆还比较滞后^[3],相当一部分图书馆服务方式传统、服务内容单一、服务能力偏低。

截至 2020 年末,西安图书馆建设了 41 个分馆、12 个服务点,均已实现通借通还。区县图书馆建分馆 54 个、服务点 37 个,实现通借通还的分馆 26 个,服务点 1 个。由此可见,通过中心馆(西安图书馆)、总馆(区县图书馆)连接村镇社区(分馆、服务点),以“西安市公共图书馆集群信息化管理平台”为纽带构建的“市域公共图书馆服务联盟”初见成效。但是,西安与全国相比仍有较大差距^[4],主要表现为投入(设施、资源、队伍)和产出(效能)双重滞后(详见表 1)。

2 面向 2035 的高质量发展愿景

“提升公共文化服务水平”

是“十四五”的重要篇章,图书馆界以“文化强国”为己任,纷纷对面向 2035 公共图书馆高质量发展建言献策。

金武钢^[5]认为,公共图书馆高质量发展将推动图书馆服务从普遍均等走向优质均衡。王世伟^[6]认为,“十四五”时期公共图书馆高质量发展应指向深度一体化、深度均等化、深度智能化、深度绿色化和深度安全化。柯平^[7]等指出公共图书馆高质量发展的十个新主题,并从组织赋能和技术赋能两大视角探讨高质量发展的驱动力^[8]。方家忠^[9]认为高质量发展的实质是系统变革。杨玉麟^[10]强调要“扎实细致”解决非政府投资兴办公共图书馆准入制度等影响高质量发展的相关问题。

综上所述,公共图书馆高质

生成,检索路径的设置,服务于明确的问题意识和研究目的。其具有与特定研究领域联系紧密,信息检索结果准确以及解决问题相对彻底等特点。元数据内容的精确性、完善性和数量的丰富性,以及检索路径的多元性,是其建设要点。

参考文献

- [1] 王洪波. 历代进士登科数据库 [N]. 中华读书报: 2019, (2) .. [2] 龚延明, 祖慧. 宋代登科总录 [M]. 桂林: 广西师范大学出版社, 2015. [3] 王瑞来. 宋代科举人物的渊薮——

- 评龚延明主编《宋代登科总录》[N]. 中华读书报, 2015, (15). [4] 傅璇琮, 龚延明, 祖慧. 宋登科记考 [M]. 南京: 江苏教育出版社, 2005 年. [5] 梁晨, 李中清. 大数据、新史实与理论演进——以学籍卡材料的史料价值与研究方法为中心的讨论 [J]. 清华大学学报. 2014, (5): 104-113.