

国史档案资源知识组织范式研究

文/雷 枫

摘要：论文阐述了以“体系分类法”和“分面分类法”构建国史档案体系类表和概念分面词表的过程和方法，并提出采用基于本体的语义技术是构建能适应历史学科动态成长的知识组织范式的趋势。

关键词：档案组织；体系分类；分面分类；主题词表；国史档案

当前由于缺少科学合理高效的信息组织方法，档案信息资源面临“资源越来越丰富，检索与利用越来越困难”的普遍困境，其使用的以档案文献为单元的传统信息组织与序化方法在面对用户基于知识内涵的深层次检索需求时捉襟见肘。构建高质量档案资源的根本方法指向了知识组织。当前，比较成熟的知识组织范式有分类法和主题法，就国史档案而言，其知识组织范式相应地包括国史学科分类体系和主题词表。

1 国史学科体系分类研究

1.1 类表编制

国史学科经历了30年发展已自成体系。笔者深入学习了党和国家重要文献（如党代会、人大报告）、权威国史著作（通史和专门史）和国史教科书，分析挖掘深藏在文本中的学科知识逻辑，充分听取学界专家对国史学科概念及其关系的逻辑分析，研究澄清了国史的知识性质。以此为依据，笔者一方面全面挖掘上述文本中的学科概念，另一方面遴选《汉表》中适用概念、析取传统分类表中与国史相关的离散类目，在此基础上，对这堆离散的、杂乱无章的概念的属性、内涵和外延进行分析，划分不同的分支及从属关系，设计建立了学科体系分类表（下称类表）。为加深学科结合，这一过程吸收了领域专家的参与，进行科学性的把握、确立学科概念层次与概念关系、选择合适术语。

类表划分为三个层次，定义了六个顶级类和按一定逻辑展开的二级、三级类的层次框架结构，其中二级类39个、三级类290个，形成学科知识骨架结构。一级类选用“政治史”“经济史”“文化史”“社会史”“国防史”“外交和一国两制史”作为类名，类目下进一步析出思想、结构、制度、实践等学科面进行类目扩展。一级、二级类名一般采用标准的学理主题词或其组合，三级类名采用富有时代特色、概念清晰、单义规范的主题词。如政治史分类见表1（不完整）：

表1

一级	二级	三级
政治史	政治思想	毛泽东思想
		邓小平理论
		“三个代表”重要思想
		科学发展观
		中国特色社会主义理论体系
	政治制度	中国共产党的执政和领导体制
		人民代表大会制度
		多党合作和政治协商制度
		民族区域自治制度和民族宗教问题
		基层群众自治制度和城乡基层民主

1.2 类表的编制方法和实用性处理

类表所采用的体系分类法是依据概念划分与概括原理和基本级别律，把概括档案文献内容的类目组成层层隶属、详细列举的等级结构体系，完整性、科学性、严密性是其主要遵循。但在编写中，为尽可能析出不同扩展维度，反映学科全貌，采用了“体系分类法与分面分类法相结合”的编写方法，设立思想、制度管理、结构、实践等学理面进行概念分析，遵循从理论到实践、从抽象到具体、从一般到特殊、从宏观到微观的逻辑来展开类目。

此外，体系类表又在类目设置上兼顾整体框架的平衡和用户查阅资料的习惯，体现了“学术性与实用性相结合”的原则。比如不严格按照学科逻辑的等级原则，而将子目多、内涵丰富、档案文献资源较多的低一级类目上调突出列类。

2 国史主题词表研究

2.1 词表编制

国史档案内容所涉及的事实对象（如人物、事件等）是人们关心的另一类主题概念。为全面反映资源特征，笔者编制了国史事实主题词表和档案资源类型表，从研究对象面（如人物、事件）、时空面（区域、历史分期）、类型面揭示主题。

为获得研究对象词表，笔者参考了国史实录、国史百科、国史编年和大事记、人物名录、中央组织史、机构名录等重要文献和工具书，挖掘遴选规范的国史术语；同时筛选中共中央文件集、中国经济工业档案、外交档案史料

等所有事实对象类关键词，获取大量的标准词、同义词、别名；并汲取《汉表》及附表中与国史有关的词汇，如各国政区名称、自然地理区和行政区划名称、组织机构名称、人物和事件名称等词汇。最终，根据历史研究对象的普遍性并结合专家分析，拟定“时间”“地点”“人物”“事件”“组织”和“理念”六个核心概念，把提取的规范术语按这六个概念分面归类，并遵循一定规则排列，构成国史档案分面主题词表：《国史人物表》《国史重大事件词表》《国史组织机构词表》《地区或区域表（包括行政区划和经济地理区域）》《国史重大分期表》《国史理念词表》，其排列规则可是字顺、范畴、词簇。表2是《国史重大事件表》片段，设置了“重大事件”“子事件”两个等级。

表2

重大事件	子事件
政协一届会议召开	制定通过《共同纲领》 制定通过《中国人民政治协商会议组织法》 制定通过《中华人民共和国中央人民政府组织法》 制定通过国旗国歌方案
中华人民共和国成立	开国大典 中央人民政府委员会第一次会议举行 中央人民政府成立
西藏和平解放	昌都战役 中国政府就西藏问题照会印度政府 和平解放西藏协议签订 解放军进驻拉萨

规范的术语系统还应包括参照系统。以上词表通过设置层级关系、同义关系和参照关系，在概念之间建立了等级、等同和相关三种简单词汇关系，以改善研究对象的语义组织：（1）主题词的层级位置体现概念间等级关系，即上下位关系或称属种关系。如《事件表》中大事件是小事件的上位词。（2）表示同一主题概念的多个术语体现出概念间的等同关系或称同义关系，体现为“见”参照。一般说来，普遍公认的常用术语作为主题词，并作为索引该概念时的最优显示，其他同义词汇（如全称、别称）作为入口词。（3）相关关系用于反映有较强语义相关性的不同分面概念间关系，如事件“真理标准大讨论”与理念“解放思想”、人物“邓小平”之间的关系。

在资源类型面，由于当代史离我们最近，史料文献浩如烟海、门类繁多，通过研究国史资源文献类型和特点，借鉴权威史料学研究成果，笔者整理出了《国史档案资源类型分类表》（见表3）。该表主要从档案文献知识特征，兼顾文献内容生产者和体裁、功能等属性进行归类。

2.2 词表的编制方法和扩展性

分面词表从国史事实概念出发，采用“分面分类”的方法，将概括文献内容与事物的主题概念组成“分面（亚面）—类目”的结构体系，一个分面词表揭示一个主题，多个分面词表揭示多个主题，其词间等同、等级、相关三种语义关系一定程度上呈现了概念间关系的规范化、弱结构化。

表3

大类	子类	备注
基本档案史料	重要文献	内容生产者为党中央、全国人大，在各种中央会议和人大会议中形成的文件汇编，是当代中国历史的珍贵记录
	专题文献	内容生产者是国家机关、军事机构和组织团体，在各种政治活动和实践中，形成大量公文材料，是当代中国历史的珍贵记录
	法规条约	包括国家法、部门法、地方法、行业法和外交条约
	中央和地方报纸	
	党政工团机关刊物	
	口述纪实回忆录	作者一般为领导人或身边工作人员，是研究人物或重大事件的重要史料
	国史人物著作文集日记	如党和国家领导人文选文稿，包含形成治国理政思想的重要论述
	其他	如族谱家谱
编辑史料	国史人物传记年谱	年谱是领导人活动的详细记载，传记是纪传体的史料
	地方和行业史志	是当代中国史资料的重要组成部分，一些典型地区、典型行业的典型材料有助于对整个历史问题作深刻说明
	国史年鉴和统计资料	通过数字展现当代中国发展的轨迹和成就

在遴选分面词表时，有大量没有找到明确概念归属的词汇，对其进行数据清洗（人工过滤）、人工聚类后，还

可析取出新的事实概念分面。如“中央农村工作会议、《共同纲领》”等词汇，可分别创建国史重（下转第16页）

采集、管理和利用各个环节的规章制度,来保护居民的隐私信息,适当采取惩戒的手段,对泄露居民隐私信息的行为进行惩罚;技术服务提供方应该建立隐私保护行业自律守则,并在服务协议中明确规定,减少不必要的管理权限,确保居民隐私信息安全;第三方机构如科研机构、媒体和行业协会等,也应建立隐私保护相关规范,在利用医疗数据进行公益研究时,应该进行脱密处理等。

3.4 强化电子健康档案隐私保护管理基础

3.4.1 建立安全责任机制。医疗机构应该建立一套责任到人的安全责任机制,对医生、护士、网络安全员、医院管理者和档案管理员的职责进行分类划分,把岗位管理职责落实到人,让各个岗位的人都树立起居民隐私保护的意识。

3.4.2 完善行业监管机制。要强化各个行业的监管。行政机构应建立有效的监管制度,联合医疗机构、媒体和行业协会,以及社会力量,对行业行为进行严格监督。加强职业道德机制建设。主要是完善各个行业领域的职业道德守则和规范,开展职业道德操守方面的培训,并建立职业道德教育和职业道德评价机制。

3.5 强化电子健康档案隐私保护技术保障

3.5.1 加强访问控制,完善系统准入审核。将电子健康档案系统按照居民、医生、档案管理员、利益相关机构等不同角色进行访问权限控制。居民对自己的电子健康档案具有最高权限,但不能对内容进行删除和修改;医生的访问权限应根据疾病种类和用户的不同进行访问区域的划分;利益相关机构只有通过居民的授权或与居民达成某种共识,才可以对部分内容进行访问。

3.5.2 配置防火墙,隔绝外部攻击。电子健康档案系统需要加强防火墙和网络入侵检测技术措施建设,防火墙可以在电子健康档案系统的内、外之间构建一个保护屏障,避免和减少黑客的非法攻击或电脑病毒。网络入侵检测技术可以实时监视用户的行为,一旦发现入侵情况立即发出警告并断开入侵者与主机的联系,保护电子健康档案隐私安全。

参考文献:

[1] 孟群.区域人口健康信息化建设与发展 [M].北京:人民卫生出版社,2014.

[2] 张春颖.信息生命周期管理研究述评 [J].情报科学,2012,30(6).

[3] 刘梅,陈金华,彭晓明.社区卫生服务机构与医院实施双向转诊的意义与建议 [J].中国全科医学,2004(01):38.

[4] 居民电子健康档案建档率超七成 [EB/OL].人民网,2017-07-11.

[5] 周晓英.电子健康档案的价值认知与应用推进策略研究 [J].档案学通讯,2018(3):109.

[6] 张悦,王茜,王健.居民电子健康档案建设中存在的问题与对策, [J].中国卫生信息管理杂志,2018,12(6).

[7] 那旭,李亚子,代涛.国外个人健康信息安全与隐私保护法制建设及启示 [J].中国数字医学,2014,9(10).

[8] 徐敏,万辉,惠朝阳,等.电子健康档案隐私保护的法律 [J].解放军医院管理杂志,2014,21(9).

作者简介:宋易倩,女,广州医科大学附属脑科医院。

(上接第9页)要会议、国史文献等分面。采用这种滚雪球似的方法,随着文献学习内容的丰富,概念分面和词汇会越来越多。另外,还有一类用户关心的综合性专题或跨越各个时期的重大历史进程,如“三农问题”“对外开放”“深化改革”等问题,因横断政治经济文化多个一级类,不便从学科分类标引,可以采用设置“综合专题”分面来处理。可见,分面词表有很强大的扩展性,可随时增补新的概念分面、随时编制分面词表。

3 结语

人文历史学科本身概念歧义性大,当代史学科还在动态成长,其史学档案资源的信息组织范式不仅需要大量领域专家的参与进行概念分析的科学性把握,还要根据学科发展不断进行修订补充。尽管存在以上风险和不确定性,分类法和主题法实际上仍肯定是档案资源知识组织范式的

要义,基于分面和体系分类方法构建的国史档案资源知识体系的思路对整个史学领域档案资源的知识组织具有很好的示范作用。

参考文献

[1] 范炜.数字环境下分面分类法解读与拓展 [J].图书情报知识,2010(05):46—52.

[2] 陈标.也谈《中图法》“中华人民共和国”类目设置 [J].当代中国史研究,2006(09).

[3] 内部资料:《军事信息资源分类法》

[4] 张注洪.中华人民共和国史文献举要 [J].历史教学,1990(1):42—47.

作者简介:雷枫,女,1968年生,中国社科院当代中国研究所副研究馆员,长期从事历史档案信息资源开发和系统建设,专注信息组织方法研究。