高校数字档案馆主题词标引应用实践

南京师范大学档案馆　杨　枫

[摘　要]计算机对文件的检索过程是一个模拟人工检索的过程，只有对电子档案文件进行科学、系统的分类，才能实现高效、便捷的查询。本文研究了在档案馆数字化过程中主题词标引的重要性以及需遵循的规范，根据高校部门设置特点制定了关键词表，提出了"3W"标准，阐明了制定主题词标引的具体方法。该主题词标引的应用实践工作可为高校档案馆建设提供参考和借鉴。

[关键词]数字档案馆　主题词　标引

电子档案文件检索的基本原则是便于查找利用和便于文件保密。因此，只有对电子档案文件进行科学、系统的分类，才能实现高效、便捷的查询。通常而言，计算机对文件的检索过程也是模拟人工检索的过程，只不过是其信息处理速度可大大超过人脑对检索词的反应速度。但是，如果档案数据库内的文件杂乱无章，文件名称不规范，这样会在很大程度上制约机检速度。如果要查找网上文件，则更不能进行盲目的无序检索。在查找之前必须理清所查找文件的网络地址、系统类别、文件制发单位和文件名称等一系列检索符号和信息，否则，检索效果将大打折扣。

在数字档案工作中，标引是对数字档案的一个加工过程，是将档案文献的自然语言转换成规范化检索语言的过程。标引既是揭示档案文献内容的重要方法，又是档案存储和检索工作中的第一个环节。科学、合理、规范的标引方法可从档案内容的角度为用户提供检索途径，使用户可以全面、准确、快速地查找到自己所需的档案信息[1]。

本文以南京师范大学数字档案馆建设过程中的主题词标引为例，对相关主题词标引的理论基础、标引原则和规范、标引流程进行阐述，以期为高校档案馆建设提供参考和借鉴。

1.主题词标引的理论基础

1.1 标引深度

在数字档案馆的建设中，将数字档案内容确定合理的标引深度是实现快速建档和快速查询的前提。档案的标引深度是指对一份档案内容进行周详标引的程度，换言之，通常可以把标引一份档案的主题数量简单地表述为给予一份档案的检索标识数量。在实际档案工作中，如果给予一份档案的检索标识少，那么其标引深度就小，这被称为浅标引；如果给予一份档案的检索标识多，那么其标引深度就大，这被称为深标引[2]。由于浅标引和深标引的标识量存在差异，检索途径也不同，因此，不同标引深度下某一目标档案被查找到的机会就存在差异。主题分析是准确确定标引深度的前提，另外，影响标引深度的因素还包括：文献本身的因素、文献本身的价值量、馆藏性质及馆藏量、检索方式（计算机检索还是人工检索）、检索要求、主题词表的质量。

1.2 主题词标引

要确定高校数字档案馆标引深度，必须先进行主题分析，确定主题之后，可把确定的主题用主题词的方式整理出来。在数字档案管理系统中，有诸多检索项目，主题词是其中关键的一项内容。主题词是指在标引和检索中用以表达文献主题的规范化的词或词组。著录项目中最具有检索意义的是主题词，而著录工作的难点也在于如何准确提炼档案主题，选好主题词[3]。主题词标引的准确、全面与否，直接影响着档案信息资源的检全率和检准率。《中国档案主题词表》是当前主题词标引的重要参考资料，然而，在当前数字档案馆快速发展的情况下，《中国档案主题词表》中确定的主题词已经不能满足实际工作的需要，特别是针对高校数字档案馆的主题词标引内容更显不足。根据以往的高校数字档案馆建设实践，并结合本单位的档案特点，笔者确定了适宜于高校数字档案馆的主题词标引方案。本文中的主题词即指的是关键词，是指在标引和检索中取自文件、案卷题名或正文用以表达文献主题并具有检索意义的非规范化的词或词组[4]。

2.数字档案主题词标引规范

2.1 言简意赅

数字档案主题词一般只选择最需要、最适合档案查询，能够代表某一类档案不同查询层次范围的主题词，其他无关紧要的主题词不录入。

2.2 长期固定

避免为提高工作效率而采用简单的"望题标引"、"字面组配"等不恰当的主题词标识方法（这样往往导致主题标识不能准确反映档案内容）。为了使档案查询者实现层次多、范围广、批量大的查询，必须将档案主题词固定下来，并杜绝同一类文件多个意思的档案主题词。如果在后续工作中发现原来设置的档案主题词不够严谨和实用，可以通过批量替换法，把原来的档案主题词修改成新的档案主题词[5]。

2.3 成表成册

数字档案管理人员要经常收集、反馈标引人员和检索人员对主题词的增、删、改意见，加强对词表的管理工作，并及时修改和增补新的主题词，以便能科学地、客观地、真实地反映当代科技领域的科目、专业术语等新词语，使主题词适合标引实践的需要。

3.主题词标引流程

3.1 关键词表

笔者根据查档频率、查档内容和归档范围给每个部门制定了关键词表，各个部门在标引主题词的时候必须使用关键词表内的词汇。比如，学校每个部门每年都要写计划、总结和规章制度，有些内容需要参照历年的资料，往往其资料范围须回溯到十年以上，某些传统的标引方案会导致工作量很大。另外，在实际工作中，若在数字档案中直接检索案卷标题经常会造成漏检，其原因在于归档的时候案卷题名标记为"任务"、"汇总报告"、"规则"等等内容，这样的内容又与实际的计划、总结和规章制度等内容不符，导致相关内容根本检索不到。因此，我们对于一些出现频率比较高的词汇"计划"、"总结"、"规章制度"等直接标引好主题词，这样在计算机检索时就非常方便。馆藏档案的存储和查询工作中很多内容均可根据使用频次和重要程度进行主题词标引，这样可使档案软件系统和扫描的文件有机结合起来，避免了纸质目录中效率较低的人工检索，从而可大大提高工作效率。笔者根据高校组织结构设置规律、职能部门工作特点和档案分类特点，总结了主题词标引规律，制定了基于数字档案的关键词表。特别是，由于笔者所在馆内平时查档频率最高的两类档案为学籍档案和人事档案，因此这两类档案的主题词标引具有普遍的参考意义。

学籍档案涉及的部门有学工处、教务处和继续教育学院。我部门所存的学籍档案包括录取名册，毕业名册，成绩，转学以及奖励和处分。我部门还经常需要对认证中心发过来的学历进行认证，工作量很大，所以必须在标引主题词的时候要做到清晰明确。比如关键词表中的"录取名册"，过去的已形成的档案中有的案卷题名为"入学名册"，有的为"新生名册"；比如关键词中"毕业名册"，过去形成的档案中有的案卷题名为"毕业生名单"，有的为"毕业生情况登记表"。这样的情况很多，如果将所有的案卷题名全部改掉，工作量非常大，我们可以在主题词中标引为"入学名册"和"毕业名册"的受控词汇就可以了。

人事档案主要涉及的部门有组织部和人事处。比如，在每次工资调整之前，人事部门需要查询部门职工的工作年限、任职经历和任职年限，以往的档案案卷标题存在很大差异，很多情况下都是按照纸质档案手工查找。特别是，在实行绩效工资前需对员工进行定级，还需根据职称评定的时间来计算，要查询各个定职称的档案文件。在案卷标题中经常会出现"干部情况登记表"，"××同志提升××院院长的通知"，如果直接在案卷标题中检索很难把握关键词，而且也经常检索不到，我们可以标引关键词"干部任免"。可根据评聘的专业技术职称级别，设置为"正高职称、副高职称、中级职称、初级职称"。关于职称"评"与"聘"的档案文件主题词同等对待。如《关于公布×××同志副研究馆员职务聘任的通知》和《关于公布×××同志档案专业副研究馆员职务任职资格的通知》主题词都是"副高职称"[6]。

3.2 "3W"标准

借鉴新闻学界的5个W（who，what，when，where，why）要素，在实际工作中提出了主题词标引的3个要素（who，what，where）。由于在档案著录项中已经有时间这一项，故在主题词标引中不再标引，删除了when这一要素；由于归档的内容很多都是文件和学籍材料，why这一项可以不必提炼。由此，主题词标引的3个要素即人物、事件、地点。

3.3 主题词标引流程

根据我馆归档的特点，在实际工作中形成了基本固定的主题词标引流程图（图1）。由于档案数量多，部门工作人员对工作内容相对熟悉，部门工作人员充当了兼职档案员的角色，我馆的归档档案是由各部门的兼职档案员和我馆工作人员共同完成的，归档是部门工作人员的一项重要任务。每年度进行归档时，移交单位将收集好的档案根据我馆制定档案管理细则与归档范围按照纸质档案特定格式进行整理，纸质档案形成后便在档案管理软件系统上录入编目。我馆于2008年建成了数字档案馆，使用了新的档案管理软件系统进行档案的数字化录入、基于主题词标引的电子编目和查询。录入编目的内容包括：档案号，盒号，归档单位，成文时间，责任者，案卷题名等。在数字档案馆中，主题词标引是其中一项重要内容，然而，有的档案管理人员会直接通过案卷题名查询档案，从而忽视此项内容。主题词标引是录入编目过程中花费时间较长的一项工作。当编目录入好的档案递交档案馆后，馆

内工作人员会进行如下工作：会对录入项目进行初步审核；与各职能部门兼职档案员沟通如何准确的标引主题词；指导和协助各职能部门的档案管理人员进行主题词标引的校对工作，使其主题词标引完全符合要求。

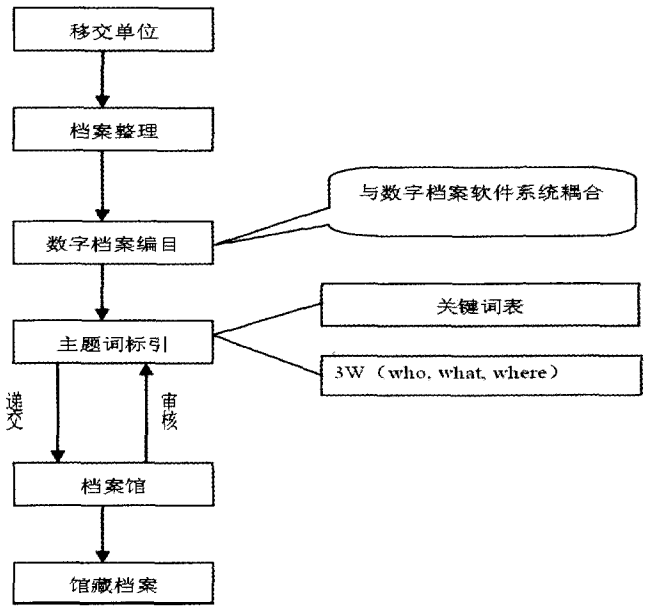

图1 数字档案的主题词标引流程图

为了使馆内工作人员和各职能部门的档案管理人员在标引主题词时有一个既合理又简单易行的参照标准，笔者根据各部门历年来归档的内容和查档使用频率制定了主题词（关键词）表，并且根据提出的"3W"标准制定标引方案。在标引主题词的时候须依照关键词表，并遵从"3W"的标准进行相关工作。以人事处的一份档案文件为例，案卷题名为"张大梅同志的任职通知"，如果只根据题名那么就只能标引"张大梅"和"任职"，根据关键词表我们要标引"干部任免"，而文件提到任职为正处级别，我们还需标引"正处"。根据"3W"标准，"who"即为"张大梅"，"what"在关键词中已经选定"干部任免"，而"where"需要标引张晓梅任职到哪，文件中指明为教务处，则需要标引"教务处"。又如，财务处的一份文件案卷题名为"上海浦东发展银行与学校贷款利息合同"，根据关键词表将该档案标引为"银行贷款"，根据"3W"标准，需要进行标引的内容为"上海浦东发展银行"，"学校"，"利息"，在整个文件中没有涉及"where"的可以不必标引。最后，对纸质文件进行数字化扫描，纸质档案和数字档案分别归入纸质档案库和数字档案数据库[7]。

4.基于主题词的数字档案查询

在将所有数字档案的主题词标引工作完成后，其查询工作就较为快捷。由于数字档案系统的查询以计算机运算为载体，速度很快，可实现深度的精细查询。如图2所见，基于主题词的数字档案查询流程包括：查询分类、确定主题词（关键词）、数字档案定位、数字档案验证。查询的结果不仅包括数字档案的存储位置、存储形式，还可确定纸质档案的准确存放位置。在实际工作中，基于主题词的数字档案查询流程比单纯的纸质档案检索过程平均效率提高30%以上，可大大节省工作时间。

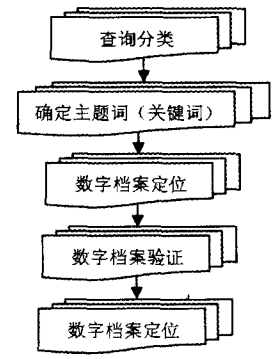

图2 基于主题词的数字档案查询流程图

5.展望

主题词标引工作对于数字档案的计算机检索尤为重要。主题词标引是一项技术性强、过程繁琐的工作，标引工作质量的好坏，直接影响档案存储的效率和自动化检索的质量，关系到数字馆藏档案作用的发挥。结合高校档案的特点，对高校数字档案馆馆藏档案的主题词标引不断完善和探索高效的主题词标引方案是档案工作者一项重要的工作。今后的主题词标引工作须根据以往的计算机检索记录的主题词（关键词）出现频次优化关键词表，并反馈于数字档案编目系统中，以实现更加高效和准确的数字档案馆主题词标引流程。

参考文献

[1]肖红琳.浅议档案的标引深度[J].2003(10):34-35.
[2]胡胜友.标引深度新探[J].档案,2000(5):34-35.
[3]江仙菊.浅议档案主题词的设置革新与利用[J].福建论坛(社科教育版),2009(8):71-72.
[4]赵芳.《档案著录规则》应用中的局限性分析[J].兰台世界,2009,(16):13-14.
[5]刘启恕.主题词的表述及规范[J].武汉金融高等专科学校学报,1995,(02):61-62.
[6]李和平,李超,刘波.档案检索系统中的多级著录研究[J].兰台世界,2009,(10):11-12.
[7]王光玉.档案征集工作规范化管理初探[J].档案学研究,2009,(02):20-23.

（上接第443页）

$$u_{rel}(m)=\sqrt{u_{rel}(C_N)^2+u_{rel}(f)^2+u_{rel}(A_1)^2+u_{rel}(A_2)^2+u_{rel}(A_3)^2+u_{rel}(R)^2+u_{rel}(V_b)^2}=0.033$$

$$u_{rel}(V_0)=\sqrt{u_{rel}(L)^2+u_{rel}(t)^2+u_{rel}(P)^2}=0.012$$

$$u_{rel}(C)=\sqrt{u_{rel}(m)^2+u_{rel}(V_0)^2}=0.035=\frac{u(C)}{\bar{C}}$$

（大气样品平均浓度为 $\bar{C}=0.434mg/m^3$）

$$u(c)=u_{rel}(C)\bar{C}=0.035\times 0.434=0.015mg/m^3$$

5.2扩展不确定度U

取包含因子k=2，则 $UV_0=k\cdot u(c)=2\times 0.015=0.030mg/m^3$

6.结论

应用纳氏试剂分光光度法GB/T18204.25-2000测定公共场所空气中的氨，本法测量不确定度主要来源为标准溶液、标准曲线拟合、分光光度计、样品重复测定、吸收液体积和采集大气样气体标干体积这六部分引入的不确定度。其中采样部分（V_0）引起的相对不确定度为0.012；由于吸收液中待测物质量(m)引起的相对不确定度为0.033。由表4可见最大的相对不确定度分量，是样品重复测定引入不确定度为0.027；其次是标准曲线拟合引起的相对不确定度为0.012。本次测量结果为：$0.434\pm0.030mg/m^3$，k=2。

参考文献

[1]国家质量技术监督局.JJF1059-1999测量不确定度评定与评定表示指南[S].北京：中国计量出版社，2001
[2]公共场所空气中氨的测定纳氏试剂分光光度法GB/T18204.25-2000[S]
[3]JJG956-2000大气采样器检定规程[S]
[4]国家环保总局.HJ/T167-2004室内环境空气质量监测技术规范[S]

（上接第444页） NSC810，NSC810利用计数器获取规定时间内的脉冲数并通过NSC800送往地面主机，地面主机通过分析脉冲数与具体数值间的对应关系确定参数的数值。

3.结论

本文对MWD系统的信号采集与处理过程进行了系统的分析和讨论，针对其中的滤波过程采用了滑动平均滤波方法进行实验仿真，结果表明对于解决现有MWD系统中滤波效果差，检测时间长的问题有明显的改善作用，对于今后国产MWD仪器的研制和改进有较好的借鉴作用。

参考文献

[1]张涛,鄢泰宁.无线随钻测量系统的工作原理与应用现状[J].西部探矿工程,2005,2
[2]盛明仁,王振光,李军成.LWD测量系统在桩12平5井中的应用[J].石油钻探技术,2000,28(3)
[3]逄玉叶.无线随钻测量系统中信号处理[J].电子测量技术,2005,2
[4]韩泽西.工程测量中测量次数的选取[J].石油仪器,1997,11(3)
[5]鄢泰宁,张涛.检测技术及钻井仪表[M].武汉：中国地质大学出版社,2009