

●张正强

基于XML电子文件管理元数据主题 元素的标引研究*

摘 要 在电子文件管理元数据中,主题元素存在着3种语义结构:无级次语义结构,有级次语义结构,多级次语义结构。由此也就形成3种不同的XML语法结构。基于XML电子文件管理元数据的主题元素的语义结构设计,不能忽视XML检索功能的需求分析。表8。参考文献5。

关键词 电子文件管理 元数据 主题元素 语义结构

分类号 G254

ABSTRACT In the metadata of electronic file management, there are three semantic structures, and then three different XML grammatical structures. The design of semantic structures based on the subject elements of XML electronic file management metadata should not ignore the analysis of the requirements of XML retrieval functions. 8 tabs. 5 refs.

KEY WORDS Electronic file management. Metadata. Subject elements. Semantic structure.

CLASS NUMBER G254

1 研究基于XML电子文件管理元数据主题元素标引的意义

在主题标引领域,基于XML(可扩展置标语言)电子文件管理元数据主题元素的标引,是一个新兴的研究领域。它既有一般主题标引的共性,又具有电子文件管理所需的特性。就其共性而言,它是一种揭示电子文件主题内容的语义标引,而就其特性而言,它又是一种基于XML语法的置标标引。

在基于XML电子文件管理元数据中,主题标引是通过主题元素来实现的。这里所说的“主题元素”,特指元数据中的主题元素;而所说的“主题标引”就是以主题标识揭示电子文件内容特征的过程。这种标引涉及两个方面,一是语义表达,即概念标引;二是语法表达,即符号标引。概念标引涉及主题元素的语义结构,符号标引涉及主题元素的XML语法结构。研究基于XML电子文件管理元数据主题元素的标引,必须从这两个方面着手。

研究基于XML电子文件管理元数据主题元素的标引,其意义在于:发现主题元素语义结构的类型;发现主题元素语义结构与语义表达、XML语法表达之间的关系;发现主题元素语义结构与XML检索功能

之间的相互作用。在这基础上,为我国建立基于XML电子文件管理元数据的主题元素的语义结构和语法结构找出规律性的认识,这对于更科学地建立我国电子文件管理元数据标准具有十分重要的指导意义。

基于XML电子文件管理元数据中主题元素的标引,其起点概念标引,是根据主题元素的语义结构进行的。而主题元素的语义结构根据其特点基本上可以分成3类:第一类是无级次语义结构,第二类是有级次语义结构,第三类是集成语义结构。相应地也就形成了3种不同的XML语法结构。

2 基于主题元素的无级次语义结构和XML语法结构的标引

在元数据标准中,主题元素的语义结构最常见的一种结构就是无级次语义结构。这种结构的特点是:作为元素的“主题”,其语义结构是无级次结构,按这种语义结构所标引的主题词也不分等级,平等地处于同一级次上表达电子文件的主题。如在著名的《都柏林核心元数据集》和美国国家标准《ANSI/NISO Z39.85—2001 都柏林核心元数据集》^[1]、国际标准《ISO 15836:2003(E) 信息和文献——都柏林核心元

* 本文为国家社科基金重点资助项目“基于XML电子文件管理元数据标准研究”系列研究成果(项目批准编号04ATQ002)。

数据集》^[2]中,主题元素的语义结构就是一个无级次的语义结构。作为上位元素的“主题”,只有一个元素,下边没有分出任何下位子元素,即无级次,所以是无级次语义结构。而在基于 XML 电子文件管理元数据中,有些标准就直接继承了《都柏林核心元数据集》中主题元素的语义结构,也就形成了相同类型的无级次语义结构。如英国文件管理署制定的基于 XML 的《电子文件管理元数据标准》^[3]中主题元素的语义结构就是如此。

对于这样无级次语义结构的主题元素,在标引时就只要直接将电子文件的主题直接揭示出来即可。如:对于《关于深化基础教育改革的规定》这一电子文件,根据无级次语义结构,在进行概念标引时就可直接标引“基础教育;教育改革”(见表 1)。

表 1 无级次语义结构的概念标引

电子文件	主 题 标 引	
《关于深化基础教育改革的规定》	概念标引	基础教育;教育改革

从表 1 可以看出,采用无级次语义结构进行概念标引时,标引的结果“主题词”,如:“基础教育”和“教育改革”都是处于同一等级级次水平上,彼此间不存在任何上下级次关系。对于无级次的语义结构,在符号标引上也表现为无级次的,下面我们就采用 XML 语法以无级次语义结构为基础,对《关于深化基础教育改革的规定》这一电子文件进行符号标引(见表 2)。

表 2 无级次语义结构的符号标引

电子文件	主 题 标 引		
《关于深化基础教育改革的规定》	概念标引	基础教育;教育改革	
	符号标引	<主题>基础教育;教育改革</主题>	
		XML 起始标识符 (下同)	XML 结束标识符 (下同)

从表 2 的符号标引可以看出,XML 语法结构显然是一种无等级级次结构。标识符号“<主题>、</主题>”对主题词“基础教育;教育改革”起着标识的作用,而被标识的“基础教育;教育改革”这两个主题词的地位处于同一等级级次上,并无任何级次之

分。主题在语义结构上是无级次的,在 XML 中表达也是无级次的。

这说明在电子文件管理元数据主题元素的标引中,主题元素的语义结构决定其概念标引的结构,而概念标引的结构又决定符号标引的结构。

3 基于主题元素的有级次语义结构和 XML 语法结构的标引

在基于 XML 电子文件管理元数据中,主题元素的有级次语义结构是指:主题元素在语义上是有着等级级次的结构,按照这种语义结构所标引的主题词也分等级级次,并在不同级次上一级比一级更专指地表达电子文件的主题。如在国际电子文件元数据领域中,澳大利亚维多利亚州文件管理署制定的基于 XML 的《电子文件管理元数据标准》^[4]是最早制定的基于 XML 的元数据标准,它的主题元素采用的就是有级次语义结构,其语义结构如表 3。

表 3 澳大利亚维多利亚州基于 XML 的《电子文件管理元数据标准》主题元素有级次语义结构

上位元素	下位子元素
主题	级次 1
	主题词——主题对象
	级次 2
	主题词——专指主题
	级次 3
	主题词——专指方面

从表 3 中可以看出,在该标准中主题元素下的下位子元素有 3 个主题词级次,在每一级次下才是相应级次的主题词。“级次 1”及其“主题词”是用来标引电子文件的主题对象的,“级次 2”及其“主题词”是用来标引电子文件的专指主题的,“级次 3”及其“主题词”是用来标引电子文件主题的专指方面的。这种语义结构显然是一种有级次语义结构。根据这种语义结构进行主题的概念标引,就可以形成由泛指到专指的等级级次分明的主题。如对《关于推广水稻旱地育苗栽培技术报告》这一电子文件进行主题的概念标引,根据有级次语义结构,就可逐级地揭示该电子文件的各级主题(见表 4)。

表 4 有级次语义结构的概念标引

电子文件	主 题 标 引	
《关于推广水稻旱地育苗栽培技术报告》	概念标引	级次 1
		主题词:水稻
		级次 2
		主题词:旱地育苗
		级次 3
		主题词:稀植栽培

采用有级次语义结构进行概念标引时,标引的结果“主题词”是处于不同等级级次上的,如:“级次 1”下表示“主题对象”的主题词“水稻”,处于最上级次;“级次 2”下表示“专指主题”的主题词“旱地育苗”,处于第二级次;“级次 3”下表示专指方面的主题词“稀植栽培”,处于最低级次。由此构成了主题语义上不同的等级级次关系。对于有级次的语义结构,在符号标引上也表现出很明显的等级级次关系。下面采用 XML 语法,以有级次语义结构为基础对《关于推广水稻旱地育苗栽培技术报告》这一电子文件进行符号标引(见表 5)。

表 5 有级次语义结构的符号标引

电子文件	主 题 标 引	
《关于推广水稻旱地育苗栽培技术报告》	概念标引	级次 1
		水稻
		级次 2
		旱地育苗
		级次 3
		稀植栽培
	符号标引	<主题>
		<级次>1</级次>
		<主题词>水稻</主题词>
		<级次>2</级次>
		<主题词>旱地育苗</主题词>
		<级次>3</级次>
		<主题词>稀植栽培</主题词>
		</主题>

在表 5 中,XML 语法结构呈现出明显的等级结构性。首先,XML 标识符号“<主题>、</主题>”对主题词起着标识的作用;其次,被标识的各个主题词在语义上处于不同的等级级次。如“<级次>1

</级次>”下所标识的主题词是属于主题对象的部分,处于最高级次,“<级次>2</级次>”下所标识的主题词是属于专指主题的部分,处于第二级次,而“<级次>3</级次>”下所标识的主题词是属于主题的专指方面的部分,处于最低级次。第三,XML 标识符号标识的主题词及其级次,形成了一个由泛指主题到专指主题的等级结构。

这说明在电子文件管理元数据的主题标引中,主题元素语义结构的等级级次决定其概念标引的等级级次,而概念标引的等级级次,又决定着符号标引的等级级次。

4 基于主题元素的集成语义结构和 XML 语法结构的标引

在基于 XML 电子文件管理元数据中,主题元素的集成语义结构是指:主题元素中既有分类的语义成分,又有主题的语义成分,两者通过集成所形成的主题语义结构。按这种语义结构所标引的主题也由两部分构成:分类部分和主题部分。从而形成了既有从分类角度表达的类主题,又有从事物角度表达的专指主题。如欧盟在调查了英国、法国、爱尔兰、芬兰、冰岛、瑞典等 10 个成员国元数据标准后所制定的基于 XML 的《电子政务元数据模型》^[5]中,主题元素采用的就是集成语义结构,其语义结构如表 6。

表 6 欧盟基于 XML 的《电子政务元数据标准》
主题元素集成语义结构

上位元素	下位子元素
主题	所属类别
	主题词
	所属课题
	所属项目

在欧盟的这个标准中,主题元素下有 4 个下位子元素,即“所属类别、主题词、所属课题、所属项目”。“所属类别、所属课题、所属项目”这 3 个子元素都是从不同角度对主题所进行的分类,即分入所属的类别、分入所属的课题、分入所属的项目。而“主题词”则是从事物角度对主题的揭示。这种语义结构很明确地表明了这是一种分类主题集成的语义结构。根据这种语义结构进行主题的概念标引,就可以达到既揭示主题类别的分类功能,又可以达到揭示事物主题

的专指功能。如对《关于冠心病预防计划》这一电子文件进行主题的概念标引,根据集成语义结构,就可以从分类与主题多个角度集成地加以揭示(如表7)。

表7 集成语义结构的概念标引

电子文件	主 题 标 引	
《关于冠心病预防治疗计划》	概念标引	所属类别:心脏病
		主题词:冠心病;预防;治疗
		所属课题:公民健康计划
		所属项目:冠心病

采用集成语义结构进行概念标引,标引的结果“心脏病”,这一词是表示《关于冠心病预防治疗计划》电子文件主题内容所属类别的,是一种分类标引,只不过这种分类标引所标的不是分类号,而是语词;“冠心病;预防;治疗”是表示该电子文件主题内容的专指主题词;“公民健康计划”一词是表示该电子文件的研究内容所属的是什么课题的;而“冠心病”这一词是表示该电子文件的研究内容所属的是什么项目的。整个主题的结构是一个分类与主题集成的语义结构。对于集成语义结构,在符号标引上也表现得很清楚。下面我们就采用 XML 语法,以集成语义结构为基础对《关于冠心病预防治疗计划》这一电子文件进行符号标引(见表8)。

表8 集成语义结构的符号标引

电子文件	主 题 标 引	
《关于冠心病预防治疗计划》	概念标引	所属类别:心脏病
		主题词:冠心病;预防;治疗
		所属课题:公民健康计划
		所属项目:冠心病
	符号标引	<主题>
		<所属类别>心脏病</所属类别>
		<主题词>冠心病;预防;治疗</主题词>
		<所属课题>公民健康计划</所属课题>
		<所属项目>冠心病(</所属项目>
		</主题>

XML 语法结构具有集成性,它集成了分类置标和主题置标,即在 XML 标识符号“<主题>、</主题>”之间是被 XML 标识符号集成的“<所属类别>、

<主题词>、<所属课题>、<所属项目>”,由此,该电子文件主题的“所属类别、主题词、所属课题、所属项目”被表达得清清楚楚。由于主题在语义结构上是集成的,所以在 XML 中表达也是集成的。

这说明在电子文件管理元数据的主题标引中,主题元素语义结构的集成性决定其概念标引的集成性,而概念标引的集成性,又决定着符号标引的集成性。

5 主题元素标引中语义结构与 XML 检索功能的关系

通过上面的分析可以发现,在电子文件管理元数据主题元素的标引中,语义结构对概念标引和符号标引起着规定性的作用,但是,语义结构归根到底是要实现检索的功能,而这种功能的实现却又要通过 XML 的语法结构来实现,所以,XML 的检索功能又对语义结构具有反作用。根据系统论的原理,结构决定功能,功能反作用于结构,这样我们就可以从结构决定功能角度得到如下规律性认识:语义结构的无级次性决定 XML 检索功能的单一性;语义结构的有级次性决定 XML 检索功能的灵活性;语义结构的集成性决定 XML 检索功能的综合性。

同样,由功能反作用于结构角度,还可以得到如下规律性的认识:XML 检索功能,其需求的一般性决定着主题元素的语义结构简单性;其需求的多样性决定着主题元素语义结构的专指性;其需求的多角度性决定着主题元素语义结构的丰富性。

在此基础上,就可以明确:

(1)主题元素的无级次语义结构,其 XML 检索功能单一,是源于检索功能的需求的一般性所决定的。因此,这种结构具有跨领域、跨专业的广泛的适应性。但是也由于此,其检索功能的专业特色不明显。

(2)主题元素的有级次语义结构,其 XML 检索功能灵活,是源于检索功能的需求的多样性。也就是这个原因,这种结构具有很好的泛指性检索、专指性检索和组配检索的优势,能够适应电子文件事物主题的层次性与专指性的要求,但是其检索功能的角度单一,不具有反映主题的类属功能。

(3)主题元素的集成语义结构,其 XML 检索功能丰富,是源于检索功能的需求的多角度性。也就是这个原因,这种结构既可以实现分类检索,又可以实现主题检索;既可以多途径检索,又可能满足不同用户、不同需求对电子文件主题不同检索的要求,因而能够充分反映电子文件管理领域的专业特色。(转第 106 页)

4.2 应用性探索缺乏合作

高校信息素质教育的应用性探索领域的工作不单是高校图书馆一个单位的工作,还需要图书馆员和学科教师、电教人员、教育学家合作。

目前,与学科教师合作进行信息素质教学已得到高校图书馆员的广泛重视,并通过学科馆员制度努力开展与学科教师的合作。但与电教人员、教育学家的合作,在大多数高校图书馆没有得到重视。

除了不同学科领域之间的合作,应用性探索领域的合作还包括高校图书馆之间的合作。高校图书馆之间有许多共性,可以共建共享信息素质网络教育资源、信息素质评价工具等许多资源,既可避免资源的重复开发建设,又能利用团体的力量,将资源的建设水平提升到一个新的高度。但目前高校图书馆界缺乏这样的合作。

参考文献

- 1 Christina S. Doyle. Information literacy in an information society. ERIC Digest. <http://www.ericdigest.org/1995-1/information.htm>(2005-02-20 查询)
- 2 Charles R. McClue. Network literacy: a role for libraries? Information Technology and libraries, June 1994,13,2:115-125
- 3 Christine Bruce. Seven faces of information literacy in higher education. <http://sky.fit.qut.edu.au/~bruce/infolit/faces/faces1.htm>(2005-02-20 查询)
- 4 Introduction to information literacy. <http://www.ala.org/ala/acrl/acrlissues/acrlinfolit/infolitoverview/introinfolit/introinfolit.htm>(2005-02-20 查询)
- 5 许祖茂,张桂花. 高等教育评估理论与方法. 北京:中国铁道出版社,1997.
- 6,15 Ann Grafstein. A discipline-based approach to information literacy. The Journal of Academic Librarianship. 2002, 28(4)

- 7,10 Information skills in higher education: a SCONUL Position Paper. http://www.sconul.ac.uk/pubs_stats/pubs/99104Rev1.doc(2003-12-12 查询)
- 8 Information Literacy Competency Standards for Higher Education. http://www.ala.org/Content/NavigationMenu/ACRL/Standards_and_Guidelines/Information_Literacy_Competency_Standards_for_Higher_Education.htm(2003-09-07 查询)
- 9 Information literacy publications. <http://www.anu.edu.au/caul/info-literacy/publications.html>(2005-02-20 查询)
- 11 Karen Macpherson. An information processing model of undergraduate electronic database information retrieval. Journal of The American Society for Information Science and Technology, 2004,55(4)
- 12 庄道明. 信息素养融入大专院校课程之探讨——以“资料搜集与报告写作”课程为例. 台北,资讯素养与终身学习国际研讨会,1998
- 13 曾晓牧,孙平. 高校图书馆的信息素质教学体系. 见:第二届国际图书馆论坛论文集. 上海:上海科学技术文献出版社,2004.
- 14 Patricia Iannuzzi. We are teaching, but are they learning: accountability, productivity, and assessment. Journal of Academic Librarianship, 1999,25(4)
- 16 Lisa G, O'Connor. Applying systems design and item response theory to the problem of measuring information literacy skills. College and Research Libraries,2002,63(6)

曾晓牧 硕士,清华大学图书馆馆员。通信地址:北京。邮编100084。

孙平 清华大学图书馆研究馆员。通信地址同上。
(来稿时间:2005-06-07)

(接第59页)

基于以上分析与研究,可以得出如下结论:

(1)基于XML电子文件管理元数据的主题元素的语义结构设计,决不能忽视XML检索功能的需求分析,这是关键,即要体现形式(语义结构)与内容(XML检索功能)的统一。

(2)对XML检索功能的需求分析,既要能体现多种类的检索需求,又要能体现多角度的检索需求,即要体现质和量的统一。

(3)需求分析要采用多种方法,既要采用文献保证法、理论演绎法,又要采用访谈归纳法、模型分析法,以保证需求分析的准确性、科学性和实用性,即要体现理论与实践的统一。

参考文献

- 1 American National Standard (ANS)/National Information

Standards Organization(NISO). The Dublin Core Metadata Element Set,September 10,2001

- 2 International Standard Organization (ISO). Information and Documentation——The Dublin Core Metadata Element Set, Feb 26,2003
- 3 Public Record Office Standard of UK. Requirements for Electronic Records Management Systems——Metadata Standard,final version,2002
- 4 Public Record Office Standard of Victoria, Australia. VERS Metadata Scheme——Specification 2,July 31,2003
- 5 European Union. Metadata model based on the recommendations of the national administrations, January 13,2003

张正强 教授。通信地址:上海四平路2575号南京政治学院上海分院信息管理系统。邮编200433。

(来稿时间:2005-05-08)