

●张 茜

电子文件管理元数据词类结构及 词类结构体系研究*

摘 要 电子文件管理元数据作为描述资源的工具正越来越多地被电子文件管理工作所使用。电子文件管理元数据是一种语言,同其他语言一样,在词汇的层面上具有词的类别。电子文件管理元数据词类结构体系可以分为基本体系与扩展体系。充分认识电子文件管理元数据词类结构和词类结构体系,对于建立我国电子文件管理元数据标准具有重要指导意义。表9。参考文献4。

关键词 元数据 词类结构 词类结构体系 电子文件管理

分类号 G254

ABSTRACT As a resource description tool, electronic file management metadata is widely used in electronic file management. As a kind of languages like other languages, electronic file management metadata have word types at the level of vocabulary. The morphology structure system of electronic file management metadata can be divided into basic system and extended system. The author thinks that it is very important to sufficiently recognize the morphology structure and morphology structure system. 9 tabs. 4 refs.

KEY WORDS metadata. Morphology structure. Morphology structure system. Electronic file management.

CLASS NUMBER G254

1 电子文件管理元数据研究的视角

电子文件管理元数据的研究有不同视角,其中大量的是从管理角度进行研究的。要使对于电子文件管理元数据的研究更为全面和科学,还必须从元数据语言这一视角出发进行研究,其中重要的是关于元数据词法的研究。

元数据是一种语言,关于这一点,著名的都柏林核心元数据在其介绍中专门指出过,并制定有都柏林核心元数据的语法规则。国际著名的元数据专家加拿大的汤姆斯·贝克也曾多次指出元数据是一种形式的语言,同时在其文章《都柏林核心元数据的语法》中,明确指出“都柏林核心元数据是一种语言”^[1]。元数据是一种语言,并且是一种人工控制语言,这一点是不容置疑的。

作为电子文件管理元数据,在2001年国际标准化组织颁布的《ISO 23081-1 信息与文件-文件管理过程-文件元数据 第一部分:原则》^[2]中将其定义为“自始至终地描述文件背景信息、内容和结构及其管理的数据。”这是从文件管理视角出发对元数据进行

定义。但是,从另外一个视角来看,电子文件管理元数据也是一种语言,因此有必要从语言角度对电子文件管理元数据进行研究。

作为一种语言,电子文件管理元数据是由词作为其最基本的单位构成的,而有词必定有词法。在自然语言中,词法主要是研究词的构成法则、词类结构和用法,而作为人工控制语言——电子文件管理元数据的词法,主要研究的是元数据词的构成法则、元数据词类的结构和元数据词的使用法则。但是,电子文件管理元数据词法的研究与一般自然语言词法研究有不同,有其特殊性,这些特殊性的主要方面表现在电子文件管理元数据词类结构及其词类结构体系上。

2 电子文件管理元数据的词类结构

电子文件管理元数据,作为一种语言,同其他语言一样,在词汇的层面上具有词的类别。这些词的类别,一般地分为基本词与限定词两大类。

* 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》(项目批准编号:04ATQ002)的研究成果之一。

2.1 电子文件管理元数据的基本词的结构

2.1.1 元素基本词

元素基本词是不可作为其他元素的限定词的元素词。如以1999年5月由澳大利亚国家档案局颁布的《澳大利亚联邦机构电子文件元数据标准》^[3]为例,该标准列出20个元素基本词(见表1)。

表1 《澳大利亚联邦机构电子文件元数据标准》
元素基本词

1. 责任者	11. 类型
2. 权限管理	12. 行政级别
3. 题名	13. 格式
4. 主题	14. 文件标识
5. 描述	15. 管理历史
6. 语言	16. 利用历史
7. 关系	17. 保存历史
8. 范围	18. 方位
9. 功能	19. 处置
10. 日期	20. 法规

表1中所列出的20个元素基本词是在电子文件元数据标准中已作为规范的元素而加以规定的,所以,也是在该电子文件元数据标准中首先必须使用的元素。很显然,这些元素也就成了该标准的元素基本词。

电子文件管理元数据的元素基本词,其功能主要是描述电子文件的方面特性,保证电子文件方面特性描述的全面性、规范性和一致性。从表1可以看出:该标准的20个元素基本词,都从不同方面对电子文件的特性进行描述,如责任者元素描述的是对电子文件负有责任的人员方面特性,题名元素是对电子文件外在特征方面特性所进行的描述,日期元素是从电子文件所涉及的时间方面特性所进行的描述,等等。由此可见,对一部电子文件管理元数据标准而言,元素基本词越多,对电子文件描述的方面特性就越多,所能描述的电子文件的方面特性也就越全面。

2.1.2 值基本词

值基本词,就是用于表述元素值的基本词,它反映的是元素基本词的内容,如表述“密级”这一元素基本词,其值基本词就可以有:秘密,机密,绝密等(见表2)。

表2 值基本词示例

元素基本词	值基本词
密级	秘密
	机密
	绝密

从表2看出,“密级”这个元素基本词,其值的内容是“秘密”、“机密”和“绝密”,这也是在电子文件元数据标准中已作为规范的值而加以规定的,也就是说,在描述“密级”这个元素基本词时,必须使用的值,就是“秘密”、“机密”和“绝密”。在这种情况下,“秘密”、“机密”和“绝密”构成了密级这个元素的三个值基本词。

电子文件管理元数据的值基本词,其功能主要是确定电子文件管理元数据元素基本词的内容,保证元素基本词对电子文件方面特性在内容描述上的准确性、规范性和一致性。显然,值基本词越多,对电子文件描述的方面特性的准确性就越高。

2.2 电子文件管理元数据限定词的结构

限定词,又称修饰词。限定词分为两类:元素限定词与值限定词。限定词是对元数据基本词的限定,或提供更为专指的语义信息,或提供基本词值的解释。

2.2.1 元素限定词

元素限定词是对元素基本词在内容范围上加以限定从而使其语义更为专指的词。这类限定词缩小了元素基本词的含义范围,使元素基本词对电子文件方面特性的描述更具体、更专指。以元素基本词“时间”为例,对其进行限定,就形成时间限定词,如“创建时间”、“处置时间”及“注册时间”等,如表3所示。

表3 元素限定词示例表

元素基本词	元素限定词
时间	创建时间
	处置时间
	注册时间

从表3可以看出,元素基本词“时间”的内容范围十分泛指而不具体,但是经限定词限定之后,“时间”就变得很具体,如经“创建时间”限定之后,“时间”就专指为电子文件所被创建的时间;如经“处置时间”限定之后,“时间”就专指为电子文件所被处置的时间,等等。可见限定词不改变基本词的含义范围,只是使基本词的含义范围更专指而已。同时看出,专指词与基本词存在着由专指词指向基本词的这

种“向上兼容”的性质。

元素限定词对元素基本词的限定可以有一级限定,也可以多级限定。如表2所示的限定词就属于一级限定。所谓一级限定,就是元素限定词对元素基本词进行直接限定。除了一级限定之外,还可有多级限定。多级限定是为了达到将基本词范围限定到更为专指的特定范围而设置的。如以《澳大利亚维多利亚州电子文件管理元数据标准》^[4]中的“保存历史”这一元素基本词为例,它就有多级限定,如表4所示。

表4 “保存历史”的多级限定示例

元素基本词	一级元素限定词	二级元素限定词
保存历史	保存处置	
		保存处置时间
		保存处置类型
		保存处置描述
	下次保存处置	
	下次保存处置时间	

从表4看出,元素基本词“保存历史”,它不仅具有“保存处置”、“下次保存处置”、“下次保存处置时间”这三个一级限定的限定词,还具有“保存处置时间”、“保存处置类型”和“保存处置描述”这三个二级限定词,这三个二级限定词在一级限定词“保存处置”的基础上又对元素基本词进行了再限定,从而使元素基本词“保存历史”的范围通过二级限定,由“保存历史”这一专指范围过渡到了更为专指的“保存处置时间”、“保存处置类型”和“保存处置描述”上。由此可见,多级限定的建立是根据所需描述元素基本词专指范围的大小来决定的,所需元素基本词专指范围越小,越具体,其所要求的限定级次就越多,反之则少。所以,对限定词级数的要求没有统一的规定,多级元素限定词多的可以有七八级限定。

由此可知,元素限定词的功能主要是保证电子文件管理元数据在内容范围描述上的专指性、规范性和一致性。

2.2.2 值限定词

值限定词,是对元素基本词的值在理解空间上加以限定的词。这类限定词由于对值在语法构成或理解的上下文加以了限定,所以,使值的语法构成和上下文语境更为具体。以时间为例,其值既可以用年、月、日——YYYY-MM-DD的语法格式表达,也可以用月、日、年——MM-DD-YYYY的语法格式表

达,还可以用日、月、年——DD-MM-YYYY的语法格式来表达,如果没有时间值的限定词,那么对时间值的理解空间就很大,就会造成理解上的多义性和不一致性。若用语法限定词——“ISO8601”标准来限定时间值的语法格式,将其限定为年、月、日——YYYY-MM-DD的语法格式,就会使时间值的理解空间更专指,从而消除理解上的多义性,达到理解上的一致性,如表5所示。

表5 电子文件管理元数据值的语法限定词示例

值	值限定词(语法限定词)
时间=2001-07-18	Scheme=ISO8601

同样,元素基本词的值还可以由词汇限定词来进行限定,如元素基本词——主题词的值为“区域供热”,采用不同的主题词表其上下文语义语境是不同的,由此,对该主题词的理解空间就很大,但是,采用值的词汇限定词——中国档案主题词表,对其进行限定,就可以知道,这“区域供热”主要是指城市的集中供热,由此,理解的空间大大缩小而变为专指,从而也消除了理解上的多义性,达到理解上的一致性,如表6所示。

表6 电子文件管理元数据值的词汇限定词示例

值	值限定词(词汇限定词)
主题词=区域供热	Scheme=中国档案主题词表

由此可见,值限定词的功能主要是保证电子文件管理元数据在理解上的专指性、规范性和一致性。

3 电子文件管理元数据的词类结构体系

电子文件管理元数据词类结构体系可以分为基本体系与扩展体系。

3.1 电子文件管理元数据词类结构的基本体系

以上我们探讨了电子文件管理元数据的词类结构,实际上,电子文件管理元数据不仅有词类结构,而且形成了一个完整的词类结构体系。电子文件,作为一个被描述的对象,需要具备描述其特性的电子文件管理元数据,因此,电子文件管理元数据与电子文件是描述与被描述的关系,而描述电子文件的最基本单位就是词。

在基本词中,我们用元素基本词来描述电子文件管理元数据的方面特性,从而确定所要描述的电子文件的特性范围,用值基本词来描述电子文件方面特性

的内容,从而确定所要描述的电子文件的内容范围。由此,形成了电子文件管理元数据词类结构的基本体系。如表7所示。

表7 电子文件管理元数据词类结构的基本体系

被描述的对象	描述对象的元数据	
电子文件	基本词	
	元素基本词	值基本词

从表7看出,在电子文件管理元数据词类结构的基本体系中,基本词是描述电子文件管理元数据的最基本单位,在基本词下具体分为两类词:元素基本词与值基本词。对电子文件而言,有了基本词,就能实现对电子文件描述,所以,电子文件管理元数据词类结构的基本体系是电子文件管理元数据所必须具备的体系。

但是,仅靠电子文件管理元数据词类结构的基本体系,只能满足对电子文件一般范围的描述,不能满足对电子文件描述专指性、具体化的要求,因此在这个基本体系上还需要扩展,以满足对电子文件描述专指性、具体化的要求。

3.2 电子文件管理元数据词类结构的扩展体系

在电子文件管理元数据词类结构基本体系的基础上再扩展出限定词,可以形成电子文件管理元数据词类结构的扩展体系,限定词进一步限定了基本词描述范围,使基本词的含义更精确。电子文件管理元数据词类结构的扩展体系如下:

表8 电子文件管理元数据词类结构的扩展体系

被描述的对象	描述对象的元数据			
电子文件	基本词		限定词	
	元素基本词	元素限定词	值基本词	值限定词

从表8看出,电子文件管理元数据词类结构的扩展体系是在基本词上扩展出限定词,在元素基本词上扩展出元素限定词,在值基本词上扩展出值限定词。由于元素限定词有一级和多级之分,值限定词也有语法限定词与词汇限定词之分,因此,电子文件管理元数据词类结构的扩展体系又可进一步扩展,如表9。

表9 电子文件管理元数据词类结构的进一步扩展

被描述的对象	描述对象的元数据						
电子文件	基本词				限定词		
	元素基本词	元素限定词				值基本词	值限定词
	元素基本词	多级元素限定词				值基本词	多类值限定词
		一级限定	二级限定	三级限定		语法类限定词
		限定词1					
		限定词2	限定词1				
		限定词3	限定词2	限定词1			
		限定词4	限定词3	限定词2	限定词1		
.....	词汇类限定词			

从表9看出,电子文件管理元数据词类结构扩展体系的扩展,主要是两个方面:一个方面是基于元素限定词的纵向与横向扩展,纵向扩展形成了多级限定词,如一级限定、二级限定、三级限定等,横向扩展形成了在同一级限定内的多面限定,如限定词1、限定词2、限定词3等;一类是对值限定词的扩展,形成了多类限定词,如语法类限定词、词汇类限定词,由此,构成了一个相当完整的电子文件管理元数据词类结构的体系。应该说电子文件管理元数据就是由于有

了这样一个词类结构体系,才可能最大限度地满足电子文件管理元数据对电子文件描述的需要。

4 研究电子文件管理元数据的词类结构及其体系的意义

通过对电子文件管理元数据词类结构及词类结构体系的研究,我们得到如下几方面的认识:

(1)从质上看:①元素基本词作为描述电子文件的范围具有定向作用,元素限定词和值限定词的扩展

都不会超越基本词的范围。②元素限定词虽然不能超越元素基本词的描述范围,但能缩小基本词的描述范围,具有定位的作用。③值基本词是对元素基本词的内容表达,是必须在元素基本词所规定的内容范围内的表达。④值限定词是对值所表达的内容的限定,但这种限定只限于对基本词内容的理解,而不涉及基本词的内容范围。

(2)从量上看:①元素基本词多少主要取决于被描述对象——电子文件的需要,即取决于电子文件管理所需的功能方向的多少。②元素限定词的多少主要取决于所要达到的专指度的需要,即取决于描述电子文件精确性。③值基本词的多少主要取决于具体的领域,即取决于领域的通用性与专业性的程度。④值限定词的多少主要取决于领域标准化程度的高低,即取决于领域内可以作值限定词的标准的多寡。

(3)从关系上看:①元素词——元素基本词和元素限定词之间存在着外延与内涵反比关系,即元素基本词的外延大、内涵浅,元素限定词的外延小、内涵深。②值词——值基本词和值限定词之间存在着衍生关系,即值基本词的构造在语法和语境上衍生于值限定词。

基于以上的认识,我们进一步发现:理解电子文件管理元数据词类结构和词类结构体系,对于指导电子文件管理元数据标准制定时进行“造词”具有极大的意义。

(1)对元素基本词造词而言:①由于元素基本词在质上具有定向作用,在量上取决于电子文件管理元数据所需的功能方向的多少,因此,在造元素基本词时,对电子文件管理所需功能方向多的要多造,对所需功能方向少的要少造。②由于元素基本词主要取决于电子文件管理的功能需求,因此,在造词方法上应主要采取功能需求造词法来构造、选择元素基本词。

(2)对元素限定词造词而言:①由于元素限定词在质上具有定位作用,在量上取决于所要达到的描述

电子文件的专指度,因此,在造元素限定词时,对描述电子文件专指度要求低的要少造,对专指度要求高的要多造。②由于元素基本词和元素限定词之间存在着外延与内涵反比关系,因此在造限定词时,应采用逻辑造词法来构造、选择元素限定词。

(3)对值基本词和值限定词而言:①由于值基本词在质上是对元素基本词的内容表达,在量上取决于领域的通用性程度,因此,通用性的电子文件管理元数据,值基本词就可以少造,反之,专业性的电子文件管理元数据,值基本词就可以多造。②由于值基本词和值限定词之间存在着衍生关系,因此在造值限定词时,应采用通用性、标准化程度高的通用标准作为值限定词。

综上所述,充分认识对电子文件管理元数据词类结构及词类结构体系进行研究的意义,对于建立我国电子文件管理元数据标准具有重大的指导意义,可以使我国电子文件管理元数据标准的制定由经验走向理论,由自在走向自为。

参考文献:

- [1] Thomas Baker, A Grammar of Dublin Core [J], D-Lib Magazine, October 2000, Volume 6 Number 10, ISSN 1082-9873.
- [2] 国际标准化组织.《ISO23081-1 信息与文件-文件管理过程-文件元数据 第一部分:原则》[S]. 2001-01.
- [3] 澳大利亚国家档案局.《澳大利亚联邦机构电子文件元数据标准》[S]. 1999,5.
- [4] 澳大利亚维多利亚州文件管理署. 澳大利亚维多利亚州电子文件管理元数据标准[S]. 2003,7.

张茜 南京政治学院上海分院军事信息管理系硕士研究生。通讯地址:上海市四平路2575号南京政治学院上海分院军事信息管理系。邮编200433。

(收稿日期:2007-09-07)