

林海青

数字化图书馆的元数据体系

摘要 数字化图书馆的元数据体系是数字化图书馆的基础结构。它由外部系统和内部系统组成。两个组成部分是同构关系,使外界的数字化信息元数据内容能映射到数字化图书馆系统,数字化图书馆的馆藏信息也能转换成网络世界通用的信息格式。参考文献9。

关键词 数字化图书馆 元数据体系 设计原则

分类号 G250.76

ABSTRACT Metadata architecture is a basis for digital libraries, which includes mutually isomorphic external and internal systems. Metadata of digital information in external systems can be mapped onto the internal systems of digital libraries, and collections of digital libraries can also be transformed to formats universally used in the network world. 9 refs.

KEY WORDS Digital library. Metadata architecture. Design principle.

CLASS NUMBER G250.76

1 元数据的概念与特征

关于元数据的一个有名的陈述是,元数据是关于数据的数据(data about data)。这当然不是一个精确的元数据的定义,但勾勒出了它的一个本质特征,就是元数据是数据的抽象,是用来描述和规定数据特征、相互关系以及相应操作的数据的集合。元数据也是数据,相对普通数据而言,元数据是更高一个层面的数据。在论及数字化图书馆时,广泛讨论的元数据概念,本质上说是计算机科学中的面向对象理论与方法的继承与扩展,同时也汲取了传统图书馆信息处理技术的经验。

目前一种观点认为元数据和传统的图书馆编目体系没有区别,它主要来自图书馆界。文献[1]关于元数据的阐述可以说是这种观点的典型表述:“元数据是关于数据的数据,所以它提供了诸如著作的作者、创作的日期、与相关著作的关联等基本信息。一个公认的元数据形式是图书馆的卡片目录索引。”更有人认为,元数据和编目数据只是同一事物的不同说法而已^[2]。另一种观点强调元数据是一种关于网络资源或其他资源的机器可理解性(machine understandable)信息,并且认为机器可理解是关键^[3]。它是一种不同于图书馆编目的全新概念。这种观点主要来自图书馆界以外的领域,特别是计算机界。这两种

对元数据的认识差异源自于各自所习惯关心的侧重点的不同。文献[4]阐述了这种区别,认为信息专业人员(当然包括图书馆员)主要关心的是怎样把描述信息和把什么信息写入元数据中去。而非图书馆界则关心数据的结构,即需要建立那些数据项而不是这些数据项的内容。但是,事实上,元数据结构的确立和怎样给元数据赋值,即元数据的使用规则,是元数据不可分割的两个方面。数字化图书馆一方面是图书馆传统文献处理技术的延伸,另一方面也是现代信息技术发展的结果。所以数字化图书馆元数据建立在传统技术和现代技术两个基础之上呈现出以下基本特征:

(1)元数据首先是一种编码体系,特别是指根据某种标准来对文献中的词及其他元素进行编码,从而揭示、描述文献的这些基本元素^[5]。元数据提供了一种框架体系和方法来描述、表征数字化信息的基本特征,并通过一整套公用的编码规则,将来源各异的数字化资源归纳到一个标准的体系中。数字化图书馆可以利用这种标准的编码框架体系管理、交流、传播和组织数字化信息。

(2)元数据是用来描述数字化信息资源,特别是网络信息资源的编码体系,这导致了元数据和传统的基于印刷型文献的编目体系根本区别。数字化信息具有传统印刷型信息所不具备的基本特征和属性,是

机读型信息,必须借助计算机及其网络读写和传播;是分布式的,这些信息可能并不存储在同一个地方,而可能分布在不同的数据服务器上;具有严格的格式化特征。元数据不仅要描述数字化信息的内容特征,而且更要描述数字化信息的这些基本属性,使得数字化信息得以被有效传播、交流和利用。

(3)元数据的最为重要的特征和功能是为数字化信息资源建立一种机器可理解框架。元数据主要是为了帮助计算机系统获得并理解数字化信息的基本特征,这些基本特征包括系统特征、内容特征、权利特征诸方面。所谓计算机可理解就是指利用元数据体系,使得计算机系统可以自动辨析、分解、提取和分析归纳数字化信息资源的基本特征。

根据以上讨论,我们可以这样来界定元数据:元数据是一种用来描述数字化信息资源,特别是网络信息资源的基本特征及其相互关系,从而确保这些数字化信息资源能够被计算机及其网络系统自动辨析、分解、提取和分析归纳(即所谓机器可理解性)的一整套编码体系。

2 数字化图书馆的元数据体系及其作用

数字化图书馆的元数据体系就是数字化图书馆中所有信息描述方法,即各种元数据结构及其实现模块的总和,是数字化图书馆的基础结构。元数据体系构建了数字化图书馆的逻辑框架和基本模型,它决定了数字化图书馆的功能特征、运行模式和系统运行的总体性能。数字化图书馆的运作,无论是存取过程还是检索过程,都是以元数据为基础实现的。

元数据体系在整个数字化图书馆系统中是一个十分重要的功能结构,它是数字化图书馆系统的管理和控制层,它可以被看作一个数据地图,指引数字化图书馆的数据存取。在数字化图书馆的信息发现、信息检索和信息组织诸方面,元数据都起着十分重要的作用。

元数据在数字化图书馆中的主要作用是为分布式数据发现和检索奠定基础。数字化图书馆的元数据体系所起的作用取决于数字化图书馆的基本目标和框架。文献[6]阐述了斯坦福大学的数字化图书馆项目中,元数据体系在整个数字化图书馆中的基础性作用。在斯坦福大学数字化图书馆项目中,元数据体系主要是为了满足系统 4 个方面的需求:资源自动发现(Automated Resources Discovery);提问陈述(Query Formulation);提问自动翻译(Automated Query Trans-

lation)和结果分析(Result Analysis)。当然,每个具体项目中的元数据体系所承担功能可能是不相同的,但是由于数字化图书馆的分布性特征,决定了元数据体系的 4 个基本功能范畴:描述功能、整合功能、控制功能和代理功能。

2.1 描述功能是元数据的基本功能

和传统情报检索语言一样,元数据的描述功能有 3 个方面:(1)是描述数字化信息的基本特征,使得数字化图书馆系统能够通过元数据体系,自动搜索到数字化信息。从数字化图书馆的发展现状看,数字化图书馆首先是一个搜索引擎,它能够帮助用户发现存在于互联网络上的数字化信息。元数据体系就为数字化图书馆自动信息发现提供了可能。它能为系统自动分析处理数字化信息提供按图索骥的框架。(2)是描述提问内容,数字化图书馆本质上是一个基于互联网的文献信息检索系统,其基本原理还是将用户提问与信息特征进行匹配来实现文献信息查找功能,所以需要利用和描述文献内容特征相同的描述体系来描述提问。数字化图书馆的元数据体系为用户描述提问建立了表达的框架,元数据系统将帮助用户有效、合法地表达提问,向用户展示数字化图书馆的检索途径和方法。(3)为数字化信息组织建立了描述平台。元数据系统还规范了数字化信息的表达形式和存储结构,由于元数据体系的建立并得到越来越广泛的认同,无论是包含在数据中的元数据还是和数据分立的元数据,都已成为数字化信息的组成部分。

2.2 整合功能

是指数字化图书馆元数据体系将外部的各种元数据系统,通过建立映射、翻译等方法整合成一个元数据系统的过程。从元数据发展现状看,人们已建立了许多种元数据方案,对于不同的数据类型也有相应的元数据系统来描述。同时,由于使用不尽相同的元数据系统,各种数字化图书馆系统和网络信息检索系统的检索方法和检索界面亦不尽相同,数字化图书馆的重要功能就是提供一种接入和用户界面服务,使得用户可以通过一个数字化图书馆的一个界面,访问互联网上的其他数字化图书馆和信息库。元数据体系就承担了转换、解释各种元数据系统的功能。数字化图书馆的元数据整合是一个双向的过程,一方面将用户的提问翻译成不同的元数据系统的表达形式,从而可以分别自动查找相应的数字化图书馆系统和其他网络信息库。另一方面,元数据体系还要将用不同种类的元数据系统描述的检索结果用一种元数据系统

表达出来。

2.3 控制功能

元数据的规范控制功能包括信息内容的规范化描述、规范标引和信息评估等方面。元数据体系可以通过标准描述框架来规范化描述数字化信息,如万维网联盟(W3C)制定了《资源描述框架》来描述万维网上的数字化信息。只有通过规范化描述才能保证数字化信息在不同的系统中交换,分布式的数字化图书馆才得以实现。同时,元数据还可以存放规范化的标引信息,用一种规范化的受控语言来揭示主题,有效地组织数字信息。从数字化信息资源组织现状看,对信息的主题正向着受控和非受控两种方法并举的方向发展,特别是随着图书馆界的参与,利用受控情报检索语言来描述数字化信息越来越得到重视。

信息评估是数字化信息控制的重要方面。由于互联网的开放性,数字化信息的生产处于无序状态,在互联网上可以说无用信息和有用信息共存,有益信息和有害信息混杂。所以对数字化信息的评估就十分重要了。对网络上信息的评估也是通过元数据系统实现的。我们可以通过元数据系统来记录数字化信息的评价数据,如可以利用<META>标记来设置数字化信息的内容等级。

2.4 代理功能

在网络环境下,尽量减小无效数据的传输对于节省网络资源,提高网络传输效率具有十分重要的意义。元数据的代理功能可以有效地节省网络资源。这是因为元数据是数字化信息资源的一种描述,记录了数字化信息资源的基本特征,可以基本反映信息的概貌,同时,元数据和数据相较,其数据量要小得多,可以作为完整信息的代理。用户在使用数字化图书馆时先获得元数据的内容,可以决定是否进一步获取数据。这对于检索一些大文件,如声音、图像、视频文件尤显其对于节省网络资源的意义。元数据的代理功能对于检索非结构性的信息也具有十分重要的意义。从目前计算机科学发展看,人们还没有找到检索诸如声音、图像、视频文件等非线性、非结构性的信息的有效途径。一个常用的也是可行的方法是通过元数据将这些信息的特征记录下来,将非结构性的数据特征用结构性的元数据描述出来,使得元数据成为非结构性数据的代理,从而使得这些数据可以被很好地检索到。

3 元数据体系的外部体系

元数据体系的外部系统是数字化图书馆外部的

元数据环境,即各种独立于具体系统的、被广泛承认的、通用的元数据标准的总和。随着互联网的高速发展,建立通用的元数据系统已成为一个热点,以至于有人认为,问题不在于没有元数据,而是我们有太多的互为竞争的元数据。关于元数据系统的发展现状,文献[7]作了详尽的阐述。在众多的元数据方案中,OCLC的都柏林核心(Dublin Core),美国联邦地理图像数据委员会的《数字化地理空间元数据内容标准》(Content Standard for Digital Geospatial Metadata - CSDGM)等都已经较为成熟并得到广泛认可。

无论是简单结构的都柏林核心还是复杂结构的CSDGM,元数据系统都是具有完整结构。元数据的结构模型可以分解成标识、构成方式、句法和规范控制4个部分。

3.1 元数据的标识

是元数据系统的基本组成部分,是用来表征数据特征的各种元素的总和。该部分建立了一个数据描述框架,如都柏林核心的标识部分就是题目、建立者等15个基本元素。元数据标识的组成结构,大体可以分成3个部分:数据描述部分主要描述数据本身的特征;环境描述部分主要描述数据的环境特征,如什么样的格式就对应需要什么样的处理软件;权利描述部分主要描述数据的知识产权特征,即合法使用数字化信息的基本规则。

元数据标识的组成具有一定的格式,即各种元素的基本排列方式。英国图书馆和信息网络办公室将元数据的格式划分成3种类型:(1)简单格式是一种特定的互联网络搜索引擎专用的元数据格式,它只是描述信息的位置特征。简单格式的元数据主要由非结构化的元数据数据项组成,特别是那种从数字化信息资源中自动析取出来的数据项,这种数据项没有严格的外在语义控制,不支持对子段的检索。我们检索Yahoo时就能感受到这种格式的主要特征。当我们输入一个检索条件时,得到的检索结果是一种层次化的,指向信息地址的,不定长、线性的数据链。(2)结构化格式则是一种正在形成中的信息描述标准,它主要由结构化的数据项组成,它由完整描述数字化信息资源的必备数据项组成,如作者、创建时间等。这种元数据格式具有严格的数据结构,形成一个字段型的、结构化的数据项组织。这种元数据格式支持字段检索,如用户可以专门检索著者项、题目项等,进而支持建立在字段基础上的布尔检索。(3)富格式是一种国际标准的格式。MARC格式是一种典型的富

结构元数据格式。富格式具有严格的语义规则和完整的信息描述手段,它有严格的格式规定和详尽的字段,能够精确、完整地描述信息资源。由于结构的复杂性,富格式的元数据系统主要是面向专业人员的,只有训练有素的主业人员才能有效准确地利用富格式元数据来描述信息特征。

元数据标识集的另一个特征是标识具有等级结构,CSGDM 的标识体系就典型地体现了这个特征,如 CSGDM 具有复合元素,复合元素下又会有复合元素,这样就构成了一个等级体系。MARC 格式也是这样。对于像都柏林这样的元数据系统,也可以视为只有一个等级的等级结构。元数据标识的等级结构由信息的基本属性决定,由于信息结构的复杂性,映射在元数据标识体系中,形成了标识体系的等级特征。

3.2 元数据的两种基本构成方法:分立式构成和一体式构成

分立式元数据构成将元数据和数据分离开,形成两个单独的又相互关联的数据实体。分立式的元数据系统由于和数据实体分开,其优点十分明显:分立式元数据可以有效地代理多媒体信息等大对象数据,将结构化的字符型数据代替非结构化的、非字符型的信息,如可以用文字信息将图像的主要特征通过元数据系统描述出来,形成结构化的元数据,以便有效地检索使用。同时也有效地减少了网络传输量,提高了网络传输效率。它的一个直观的现实模型是《中国期刊网专题全文数据库》,该数据库中的全文文献以扫描图像格式存储,用户通过元数据系统检索得所需文献的基本信息,然后再下载全文图像文件,利用专用浏览器阅读全文。MARC 格式存储的书目数据也是一种分立式的元数据结构。分立式元数据系统可以有效地避免元数据扩大数据长度的问题。分立式元数据比较适用于信息收集整理机构通过元数据体系有效地揭示、组织数字化信息。一体式元数据构成是将元数据和数据实体结合在一起,形成一个整体。这种方法主要是便于信息创建者自行建立元数据,使得数字化信息资源能够有效地被诸如搜索引擎、数字化图书馆等网络信息搜寻者搜寻到。一体化元数据系统是以分布式信息存储与交换为特征的互联网络发展的产物。由于互联网的开放性,使得基于互联网的数字化信息的组织和利用变得十分困难。利用一体化的元数据,让信息作者自行描述信息特征是一个有效的解决方案。

3.3 元数据的语法规则规定了元数据表达信息的基

本方法和规则

元数据语法的核心是怎样将元数据融入互联网通用语言中去,也就是说,怎样通过互联网通用语言来表达元数据。元数据系统和通用标记语言的结合主要包括两个基点:(1)将元数据具体描述成通用标准标记语言(SGML)的文献类型定义(DTD)。(2)将元数据映射到标记语言,如 HTML 的标记中去。在都柏林核心建立初期,人们就认识到元数据语法的重要性,1996 年在英国瓦威克大学召开的第 2 次都柏林核心工作组会议的主题就是讨论都柏林核心的语法问题。OCLC 瓦威克工作组建立了一个瓦威克框架(Warwick Framework),在元数据的语法方面主要解决了用通用标记语言特别是 HTML2.0 来表达元数据。CSGDM 也定义了一个 SGML 的 DTD,同时提出了严格的表述方式^[8]。HTML 的发展为元数据的表达提供了更大、更规范的空间,HTML4.0 版的 <META> 标记更加规范,使得元数据的表达也愈加规范。关于元数据语法的一个重要发展,就是万维网联盟提出的资源描述框架(RDF)。它是使结构性元数据得以编码、交换和重用的基础结构。RDF 是 XML(扩展标记语言)的一种应用。XML 需要利用结构性的约束方法来明确传递语义。所以建立在 XML 上的资源描述将更加严格和明确。

3.4 元数据的规范控制法

本质上说,元数据系统的句法就是元数据规范化的一个组成方面,但是对元数据各种标识的赋值的规范化控制也是十分重要的。赋值的规范化控制就是指通过建立某种规范化方法控制元数据标识的取值,以保障标引和检索的一致性。规范化方法主要通过枚举取值范围实现。元数据的规范化控制主要由词表和值组成,并通过对词表的规定,引入了词表所对应的规则。

元数据的发展已从各自为政的战国时代向着标准化、结构化的方向发展。随着互联网特别是万维网的兴起,建立标准的、被人们共同遵守的元数据系统已成为人们的共识,其中元数据方案的标准化已得到普遍重视,元数据的重用和各种元数据的互换已成为元数据发展的趋势。RDF 附加地提供了一个方法来建立人类可读的和机器可处理的元数据词汇,并通过这些词汇的设计,鼓励不同信息团体的元数据表的重用和扩展。结构性约束的 RDF 推动和支持结构性元数据的统一编码和交换,从而提供了不同资源描述团体所定义的元数据包具有互换性。RDF 的提出,为元

数据的发展提供了一个规范化的框架。国际标准化组织最近成立了一个元数据工作组,研究建立元数据的全球性的标准。并颁布了相关的国际标准:数据元素的专业化和标准化(ISO 11179)。国际标准化组织的介入,将大大推动元数据标准化的进程。随着元数据系统的发展,建立各种元数据系统之间的相互转换关系和方法成为规范数字化信息,保障数字化图书馆功能实现的重要条件。建立一套元数据系统的注册登记制度,及其相应的技术实现手段和相应标准已引起人们的广泛重视。

4 元数据体系的内部结构

元数据体系的内部结构主要是数字化图书馆系统本身的元数据处理方法和体系结构,不妨称之为数字化图书馆元数据管理系统。元数据管理系统是整个数字化图书馆系统的重要组成部分,其基本功能是为数字化图书馆的运行建立基础。元数据内部结构和外部系统是同构的。元数据管理系统勾画了数字化图书馆的整体轮廓。元数据管理系统的这种地位是由数字化图书馆系统的基本结构决定的。数字化图书馆是依附于互联网的。其基本结构是客户机服务器结构。数字化图书馆本质上说是一个数据服务器,用户端上的浏览器是系统的客户端,它们之间的数据传输通过公认的协议进行。数字化图书馆的元数据管理系统成为这两者沟通的桥梁。无论数字化图书馆的数据的物理存储结构是怎样的,元数据系统建立了一个存取数据的逻辑框架,它就像一个交换器,控制着数据的存取。

元数据管理系统的组织结构及其实现方法取决于数字化图书馆系统的基本目标和体系结构。如斯坦福大学数字化图书馆是一种分布式的、异类连接的、基于代理的数字化图书馆系统,它的元数据管理系统由4部分组成:属性模型代理(Attribute model proxies)、属性模型转换器(attribute model translators)、查询代理的元数据摘要(metadata facilities for search proxies)和元数据仓库(metadata repositories)^[9]。在这里,数字化信息资源首先被分成属性和属性的值两部分,属性的集合称之为属性模型。US-MARC就是一种属性模型。属性模型代理就是基于某种属性模型基础上的数据集,是一种由属性定义的信息集合。由于斯坦福大学数字化图书馆是一种分布式的、异类连接的、基于代理的数字化图书馆系统,这就需要属性模型转换器将各种不同的属性及其属

性值相互翻译、转换。查询代理的元数据摘要则同时结构化地描述查询代理所提供可存取的数据集以及查询代理所能够提供的查询能力。如通过查询代理元数据摘要可以了解某一个查询值在数据集的某一属性中出现多少次等信息,而不是获得信息本身,这样可以帮助查询者进一步修正检索策略。它还能向查询者提供某个属性可以接受的检索方式,如是否接受关键词查询等。而元数据库则存储了其他3个部分的信息,它可以从其他3个部分获得元数据信息,也允许其他3个部分直接向一个或多个元数据库提交元数据信息,这样就形成了一个本地的元数据仓库。

数字化图书馆元数据管理系统的桥梁作用决定了它的基本结构。为了实现数字化图书馆和外界信息环境的沟通,元数据内部系统和外部系统必须是同构的。这种同构关系实际是将外部元数据系统映射到数字化图书馆的内部体系中的方法。为了建立同构关系,元数据管理系统的结构包括6个组成部分。

(1)基准元数据系统。是指某个数字化图书馆标准的元数据系统,它的作用是:首先,作为基准元数据,组织标识数字化图书馆中的数字化信息资源,揭示这些信息资源的基本特征,使之成为可检索的有序体系,无论数字化图书馆的数据是怎样按何种物理结构存储的,基准元数据系统都控制着数据的存取。其次,基准元数据将描述用户的查询提问,将用户提问规范为系统认可的标准形式。第三,基准元数据将成为数字化信息的一部分通过互联网络传递到客户端、其他数字化图书馆和网络搜索引擎等网络信息发掘工具。数字化图书馆系统的基准元数据是一个较为复杂的结构,它可以被定义成几种基本元数据类型:知识描述型元数据(Intellectual Metadata)、结构型元数据(Structural Metadata)和存取控制型元数据(Access Control Metadata)。知识描述型元数据用来描述、发现和鉴别数字化信息对象,如MARC、都柏林核心等,它主要描述数字化信息资源的主题、内容特征。结构型元数据描述数字化信息资源的内部结构。相对知识描述型元数据而言,结构型元数据更侧重于数字化信息资源的内在的形式特征,如目录、章节、段落等特征。存取控制型元数据是指用来描述数字化信息资源能够被利用的基本条件和期限,以及指示这些资源的知识产权特征和使用权限。以上3种类型元数据的划分,得到数字化图书馆界的共识,如美国国会图书馆的数字化图书馆计划和密西根大学的数字

化图书馆项目都采用了这个分类方法。除这 3 种元数据类型外,内部元数据系统还应有第 4 种类型,就是评价型元数据,主要描述和管理数据在信息评价体系中的位置。

(2)元数据字典。是一种用于各种元数据体系到系统基准元数据系统相互转换的对照表,它描述了各种元数据的基本特征,构建了各种元数据与基准元数据系统的对应关系。其基本作用是为系统的转换模块提供转换依据。各数字化图书馆系统的元数据字典结构的设计可能有所不同,但一般说来,其中的每一条记录都包含元数据名称、元数据标识名称、子标识数据集、规范控制表、对应基准元数据标识和唯一的指示键等主要字段。其中,指示键是系统中用来表示元数据字典记录的唯一的编号,它和元数据的记录构成一一对应关系。元数据字典和基准元数据系统构成了一对多的关系。

(3)数据属性集。是指数字化图书馆存储数据的属性总和。它一方面是系统的数据字典,记录了系统数据的所有特征,描述了数字化图书馆的数据结构;如在数据属性集中存放了系统数据的表名称、字段名称、类型、长度、索引键等基本属性。另一方面,它和系统的基准元数据体系建立起对应关系,将每一种数据都和相应的基准数据对应起来,从而使得基准元数据体系可以控制、代理系统数据的物理存取。数字化图书馆的数据结构不必机械地和基准元数据保持一致,如系统的基准元数据系统可以采用都柏林核心,但数据结构则可建立在 MARC 基础上,元数据管理系统可通过数据属性集将数字化图书馆的数据结构和基准元数据相对照,保障它们之间的可互换性。

(4)数字化信息源特征集。数字化图书馆每一次信息查询都可能对若干个分布在不同地方并利用互联网连接的数字化信息源的检索操作。这就需要描述各个数字化信息源的基本特征,特别是其基准元数据体系。数字化信息源特征集也是一种元数据系统,只不过它描述的对象是信息源,即信息的集合而不是每一个具体信息。数字化图书馆系统可以通过信息源特征集来确定各信息源所采用的元数据体系,将用基准元数据表达的查询式转换成个信息源所采用元数据表达式,从而决定各个信息源的检索方法并解释检索结构。

(5)转换模块。以上 4 个部分可以说是数字化图书馆元数据管理系统的静态参照表,它们记录了元数据管理系统的各种对应关系,而实现各种元数据之间

的相互转换,则是通过元数据管理系统的转换模块实现的。转换模块提供了实现各种元数据之间相互转换、翻译的方法。相对于静态的对照表,转换模块是以它们为基础的、由专门程序实现的动态的过程。

(6)维护模块。是数字化图书馆元数据管理系统中,对各种静态对照部件和动态转换过程进行管理的模块。维护模块可以对各种对照表进行添加、删除、修改等动态管理,保证元数据管理系统的可扩展性和可维护性。

虽然元数据内部结构在实际系统设计上可能形态各异,但是,这 6 个组成部分是元数据管理系统的基本功能构成,它们实现了数字化图书馆对元数据的处理过程。

5 元数据体系的设计原则

在实际设计过程中,元数据体系设计具体的方法和方案可有多种选择,例如可以将元数据字典设计成一维的单表结构,也可设计为相互关联的多表结构。但无论具体设计方案有何不同,也要和其他计算机系统一样,遵循普遍的设计原则和方法。由于元数据体系的基本作用是控制数字化图书馆系统和外界进行数据交换,因此,标准性、完备性和可扩展性应该成为所有设计原则中最重要的内容。

所谓标准性是指在数字化图书馆元数据体系的设计过程中,应该以被广泛认可和遵循的标准体系为依据。元数据体系设计的核心是基准元数据系统的确定。在选择基准元数据系统时,应该选择由权威机构或组织制订的、被广泛认可和使用的元数据方案。要考虑所选择的元数据系统是否是正式的标准,如是否是国际标准、是否是业内标准等。

完备性是指数字化图书馆的元数据体系应该能够解释和兼容大多数数字化图书馆所涉及主题领域中,获得公认的元数据方案。有太多的元数据系统方案,它们都在一定领域内被广泛使用着,甚至成为业内的标准。只有兼容这些被广泛使用的元数据系统,才能够实现数字化图书馆作为一种开放性的、分布式的、跨领域的数字化信息资源管理和检索系统的基本功能。由于数字化信息的多样性,用于描述不同类型数字化信息资源的元数据是不同的,所以,数字化图书馆系统为了能够兼容不同种类的信息类型,也必须兼容相关的元数据系统。

和元数据体系设计的完备性原则相关的是可扩展性原则。为了保证元数据体系是 (下转第 69 页)

(包括电话费和上网费),明显高于计算结果。按理用户应能获得较满意的上网速度,但 CNNIC“报告 9907”显示:用户对我国因特网最不满意的的问题中“速度太慢”的排第一。

产生这一现象的原因是:上述模型只是理论上的分析说明,要应用该理论模型分析中国因特网定价,必须结合中国实际,对模型进行修正。CNNIC“报告 9907”显示,用户平常持续上网时间最常见的时间段如表 5。可见中国因特网用户上网时间有很强的分段性,55%用户的上网时间集中在 18 00 - 24 00;而在 12 00 - 15 00 却用户很少。为避免拥塞,应采用分时间段定价原则;在基本上网资费的基础上,通过对不同时间段收取不同的上网费,均衡各时间段的上网人数,以充分利用现有网络基础设施和资源,提高 ISP 的服务质量。用价格这一杠杆加速推动因特网应用,对于信息产业乃至整个国民经济的稳定增长都具有深远的战略意义。

表 5

7 00 - 12 00	12 00 - 15 00	15 00 - 18 00	18 00 - 21 00	21 00 - 24 00	24 00 - 7 00
15 %	9 %	10 %	20 %	35 %	11 %

(来稿时间:1999-11-02)

(上接第 64 页)完备的,要求其具有扩展的可能性。元数据系统是一个发展活跃的领域,新的元数据方案会不断出现,老的方案会不断修改完善。数字化图书馆的生存环境是一个不断变化的环境,新的信息源也会层出不穷。这就要求系统能够允许将新的方案或新的信息源注册到系统中,也要求系统能够修改更新已注册的元数据方案。

参考文献

- 1 Miller, Paul. Metadata for the Masses. In: Ariadne, 5, Sept. 1996. <http://www.ariadne.ac.uk/issue5/metadata-masses/>
- 2 Caplan, Priscilla. You Call It Corn, We Call It Syntax-Independent Metadata for Document-Like Objects. The Public-Access Computer Systems Review 6, no. 4, 1995.
- 3 Bemers-Lee, Tim. Metadata Architecture. Documents, Metadata, and Links. Last edit Date: 1998/02/06 17:06:

参考文献

- 1,2 中国互联网络中心,中国 Internet 发展统计报告,1999 - 07
- 3 [美] Bruce R. Kingma. The Economics of Information. Libraries Unlimited, INC., 1996
- 4 中国首届青少年国际互联网络知识大赛组委会. 中国 Internet 知识和应用情况调查报告. 计算机世界报, 1999-03-22

晏凌 武汉大学大众传播与知识信息管理学院情报学系 98 级在职研究生,现工作于武汉大学图书馆。通讯地址:武汉大学图书馆。邮编 430072。

王应解 武汉大学大众传播与知识信息管理学院情报学系 97 级研究生。

46. <http://www.w3.org/DesignIssues/Metadata.html>
- 4 Jessica Milstead, Susan Feldman. Metadata. Online, vol. 23 Issue 1, Jan/ Feb 1999.
- 5 Dorman, David. Metadata Musings. American Libraries, Jan99, vol. 30 Issue 1, p102, 1p
- 6,9 Michelle Baldonado, Chen-Chuan K. Chang et. al. The Stanford Digital Library Metadata Architecture, International Journal of Digital Libraries, 1(2), Feb, 1997.
- 7 Jessica Milstead, Susan Feldman. Metadata, Projects and standards. Online, vol. 23 Issue 1, Jan/ Feb 1999.
- 8 <http://www.fgdc.gov/metadata/fgdc-st&001-1998.dtd>

林海青 馆员,南京大学-约翰斯·霍普金斯大学中美文化研究中心图书馆工作。通讯地址:江苏南京。邮编 210093。

(来稿时间:1999-12-14)