

基于 Tamino 和 XML 的建筑数字图书馆应用研究^{*}

A Study of Architecture Digital Library Based on Tamino and XML

李 蕾 邢春晓 冯建华 周立柱

(清华大学计算机科学与技术系 北京100084)

Abstract With the drastic development of Internet and information technology, Digital Library has become an important information and knowledge management environment. How to efficiently store and search huge digital objects, especially XML document, is one of the challenges in the field of database. In this paper, we present a novel digital library architecture based on native XML database--Tamino, and apply it to the prototype of TsingHua University Architecture Digital Library (THADL). The specification of metadata DTDs 1.1 for THADL is designed and established. Finally, we give the example of THADL resources according to DTDs, and discuss the implementation of THADL based on JSP and Tamino.

Keywords XML, Tamino, DTD, Metadata, Digital library

1. 引言

中华民族有五千多年的文明史,在漫长的历史进程中,创造了光辉灿烂的中华文化,为人类文明作出了重要贡献,在这些宝贵的文化资源和遗产中,与中华文明同样源远流长的中国建筑始终占有一席之地。特别是1930年由朱启钤、梁思成和刘敦桢建立的中国营造学社,完成了中国主要有关建筑古籍的整理和出版,在15年中对2783处古建筑进行了实地考察,留下了十分珍贵的照片、测稿、图纸、建筑古籍等资料,对中国建筑史的研究具有重大的贡献,开创了研究中国建筑之先河,奠定了整个中国建筑史学的基础。但是这些具有极高成就和价值的资源正在面临着自然残毁和人工破坏双重威胁,如何科学、合理地保护、保存、展示和传播这些中华瑰宝是我们面临的重大挑战。

以因特网为代表的信息技术革命是20世纪对人类生活影响最大的事件之一,它已经把整个地球连接成为一个有机的整体,为人类展现了跨国界、跨社会、跨文化、跨语言的信息、知识交流的无限可能性。基于因特网和多媒体技术的数字图书馆为我们进行中华文化资源的保存、组织和管理提供了一条有效的途径。数字图书馆不仅仅是一个有着信息管理工具的数字收藏的等价词,它更是一个环境:将收藏、服务和人带到一起以支持数据、信息,乃至知识的全部流程,包括从创造、传播、使用到保存的全过程。数字图书馆已经成为信息高速公路的重要信息资源,是未来社会的公共信息中心和枢纽,必将成为21世纪知识经济时代一种新的基础设施。

我们构建清华大学建筑数字图书馆^[1](THADL: TsingHua University Architecture Digital Library)的主要目的包括:

(1) 开展珍贵资源的“抢救式”的数字化标引、组织和保存工作,使这些数字资源能够基于因特网向用户提供浏览、查询和检索等信息服务。

(2) 从小型示范项目做起,在理论、方法、技术和标准上取得突破,然后逐步扩大,为构建大规模的数字图书馆的研究和应用奠定基础。

(3) 建立基于XML/RDF元数据的DTD规范^[2],提供中国建筑科学元数据的格式、描述、处理和转换的标准,为THADL提供数字资源标引和描述的依据。

(4) 基于Software AG公司的Tamino数据库,研究和开发THADL的信息服务平台,建立具有开放性、可重用性和可扩展性的原型试验系统,基于因特网为用户提供友好、主动和统一的服务。

本文将重点讨论基于XML和Tamino数据库设计和开发清华建筑数字图书馆应用。

2. 基于 Tamino 的数字图书馆体系结构

随着因特网的普及和发展,其上的数字资源正以几何级数增长。据发表在《科学》杂志1999年7月的文章《WEB信息的可访问性》估计,全球的网页超过8亿,有效数据超过9T,并且仍以每4个月翻一番的速度增长。如何科学合理地进行信息编码和交换,减少信息过载成为我们面临的主要挑战之一。目前因特网上的信息主要是基于HTML(超文本标记语言)的标记文件,主要缺点在于HTML语言只重信息表现,缺乏对信息内容、结构的描述,降低了信息处理、交换和管理的效率。1998年W3C(World Wide Web Consortium)组织提出了XML(扩充标记语言)1.0标准^[3],XML能提供用户按需自行定义标记及属性名,结构化地描述信息内容,从而使XML文件的结构可以复杂到任意程度,XML良好的数据存储格式、可扩展性、高度结构化和便于网络传输的特点,为基于XML的应用赋予了强大的功能和灵活性。因此我们基于XML和Dublin Core定义THADL的元数据的规范。在研究和分析德国Software AG公司的Tamino纯XML数据库的基础上,基于Tamino API和JSP开发了THADL信息服务原型试验平台。

2.1 Tamino 简介

Tamino(Transaction Architecture for the Management of Internet Objects)^[4]是德国Software AG公司研制的第一个原生(Native)XML文档存储和管理信息服务器,是一个支持Web的完备数据库管理系统,提供数字资源交换和应用

^{*}国家重点基础研究项目资助,项目号G199032704。李蕾 硕士生,研究方向为数字图书馆和XML元数据管理。邢春晓 博士后,副教授,研究方向为海量信息存储和管理、数字图书馆等。

的集成环境。德国 Software AG 公司在数据库领域已经有近 30 年研发树形/层次数据库的经验, 这为 Tamino 数据库的研发提供技术基础。目前我们把对于 XML 文档的存储和管理系统主要分为三种, 其特点如下:

(1) 文件系统: Microsoft NT, UNIX 文件系统

a) 查询与检索相当困难。

b) 只能针对整个文档/内容进行管理。

(2) 具有 XML 功能的关系数据库系统^[7]: DB2/XML Extender, Informix/Object Translator and Web DataBlade, Microsoft SQL Server 2000, Oracle 8i/9i

a) 通过 XML 到关系模式的映射层提供 XML 数据的存储和查询。被映射到数据库中的数据以一定的模式映射到应用的时候, 原来的 XML 元数据和结构可能会丢失。仅适合平面的结构化内容, 对 XML 文档的分解丢失了其整合性。

b) 存储和检索 XML 文档, 可以通过 XML 规范技术 (XPath, XSLT, DOM, SAX) 或是关系数据库的 SQL 来完成, 基本的存储单元是和实现相关的。

c) 需要将 XML 映射到一个规格化的关系式结构, 或将 XML 信息完整无缺地以 BLOB 存储。

d) 基于 SQL 的查询需要巨大的开销, 对 XML 文档层状

树结构只能通过复杂的连接 (join) 来进行查询。

(3) 原生 XML 数据库系统^[7]: Tamino, dbXML, PDOM, Lore, X-Hive/DB。

a) 为了存储和处理 XML 数据而设计了专用的扩展的 XML 结构, 支持 XML 文档格式的输入输出, 提供原生 XML 的存储提高了系统对 XML 文档处理的性能;

b) 数据查询是基于 XML 相关标准如 XPath, XSLT, DOM, SAX 建立的, 提供了快速高效的查询响应;

c) 系统的基本的存储单元是 XML 文档, 提供了 XML 数据表示的完整性和一致性。

通过研究和分析三种方式的存储、查询和管理性能和量化指标 (由于文章有限, 将在另文中给出), 我们认为第三种方式在处理树形或更复杂结构的 XML 文档时, 具有较高效率和性能。Tamino 是我们首选的研发平台, 同时我们正与德国 Software AG 公司合作在此基础上开发了基于 Tamino 的数字图书馆信息服务试验平台。

2.2 基于 Tamino 的数字图书馆体系结构

在 THADL 的项目中, 我们基于浏览器/服务器多层模型, 提出了一个分层的数字图书馆体系结构, 主要包括表现层、服务层和存储层。

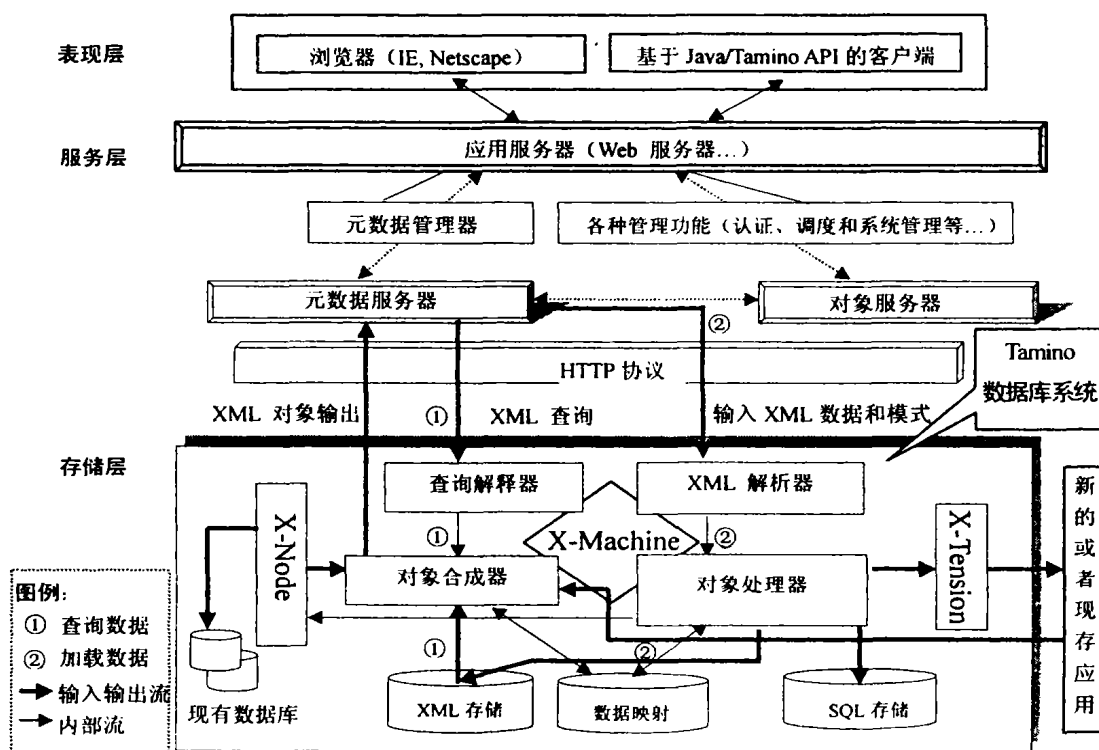


图1 基于 Tamino 的 THADL 体系结构图

表现层--基于浏览器或者是 Java 应用程序, 可以采用 B/S、C/S 给不同的客户提供统一的获取分布式的异构信息源和服务的途径。服务层--通过应用服务器提供主要的管理和服 务组件。包括检索管理、系统管理、操作管理、安全管理和元数据管理等相应的组件。通过这些组件提供相应的认证、调度和发布功能, 由元数据服务器、对象服务器和应用服务器构成的 THADL 体系将提供相应的处理和服务。存储层--它包括数据库服务器和对象服务器。数据库服务器负责管理和存储所有元数据库和基于关系数据库管理操作系统 (RDBMS) 的词典库。对象服务器提供存储机制, 并提供对 RDBMS、XML 管理系统和 NT/UNIX 文件系统的支 持。

Tamino 数据库管理系统位于存储层, 下面介绍它的几个主要的功能部件。

(1) X-Machine: 是驱动 Tamino 的引擎, 它的基本功能就是存储 XML 对象并从不同的数据源中进行查询。它是基于管理员定义在数据映射中的模式来处理存储和查询的。XML 对象存储在 Tamino 中, 但是 Tamino 也提供一个内部的 SQL 存储, 并且支持 Tamino 以外的其他数据库类型。

(2) 数据映射: 是 Tamino 的知识库。这些模式主要是由管理员制定, 包含存储和查询及组织生成 XML 对象的规则。这些模式的制定可以由图形工具来支持并保持其模式树的正确性。

(3)XML 解析器:存储在 X-Machine 的 XML 对象将由存储在 Tamino 数据映射中的模式进行描述,X-Machine 内部的 XML 解析器将会检查模式的语法正确性以确保加载的 XML 对象是“格式良好的”。

(4)对象处理器:Tamino 存储数据时,对象处理器使用数据映射中提供的相关模式信息把 XML 数据存储在 XML 存储中。

(5)查询解释器:Tamino 的查询语言是 X-Query。查询解释器分解查询请求并与对象合成器交互,按照存储在数据映射中的模式检索对象。

(6)对象合成器:使用定义在数据映射中的存储和查询规则,对象合成器能够构造信息对象并把他们作为 XML 文档返回。

基于以上功能部件,Tamino 数据库管理系统实现的主要功能是:

(1)数据的查询

Tamino 的查询语言 X-Query 基于 W3C 的 Xpath 规范,当输入查询语句时,查询解释器就会分解查询请求,按照数据映射提供的存储规则,并与对象合成器交互,组成查询结果。这时还可以通过 X-Node 把现有的数据库(例如关系数据库)中的数据连同存储在 XML 存储中的 XML 对象组织在一起并展示出来,因此对于用户来说是透明的。如上图1中①所示。

(2)数据的加载

当进行数据加载时,通过 XML 解析器来解析模式的语法正确性以确保加载的 XML 对象是“格式良好的”,然后和数据映射中提供的查询和存储的相应规则一起提交给对象处理器,如果两者一致,就可以存储在 XML 存储中。或者是通过 X-Tension 提供给一些已经存在的或是新的应用来实现整合外部的应用。如上图1中②所示。

3 基于 XML 的元数据 DTD 设计

3.1 元数据

元数据是关于数据的数据,W3C 把元数据定义为机器可

理解的网上资源的描述信息。由于网上信息量的巨大,而信息的分类又相对缺乏,没有元数据,网上的可用信息最多只是人类可理解、机器可读的,而不是机器可理解的。而信息分类的缺乏又造成了数据处理和信息检索的效率低下。Web 上现有的一些搜索引擎仍然使用传统的索引和关键字匹配的方法检索信息,检索的效率和准确率都不能令人满意。为了改善这种状况,采用元数据对信息进行分类和结构化是一个有效方法。

3.1.1 Dublin Core 简介 Dublin Core 元数据标准的开发的最初目的是描述网上资源信息。Dublin Core^[5]元数据元素集(DCMES)提供了一组语义词汇来描述信息的核心性质,如“Description”、“Creator”和“Date”等。Dublin Core 的特点是:

(1)简单性: Dublin Core 的使用对象是资源描述专家而非专业的图书馆分类和编目人员。Dublin Core 的大部分元素都很通用、易于理解,而且这些元素概括了图书馆分类卡片的复杂内容。

(2)语义互操作性: 在 Internet 上,不同的元数据的描述模型使不同学科或学科边缘的检索能力受到很大限制,采用 Dublin Core 这种通用的、可理解的描述符有助于数据内容标准的统一,增进跨学科的语义互操作性。

(3)可扩展性: Dublin Core 提供的描述模型是最简洁的、可选择的,可以增加更详细的描述,如在图书馆领域中的 MARC 分类。而且, Dublin Core 对于结构的编码和语义的丰富都有很大的灵活性和扩展性。

(4)国际影响: Dublin Core 在北美、欧洲、澳大利亚和亚洲的20多个国家中被应用于网上资源挖掘。

3.1.2 THADL 元数据 DTD 1.1 文档类型声明 (DTD)用于定义 XML 文档逻辑结构的约束,提供对预定义存储单元的支持。

THADL 项目主要针对中国建筑类的图纸、测稿、照片、汇刊、专著、论文等特色资源,采用中国图书馆分类法、汉表建筑类主题词表和规范控制等,设计和制定了 THADL DTD 1.1规范。我们以 Dublin Core 的15个元素为基本框架,将其扩

表1 THADL元数据英文DTD

<pre> <!ELEMENT rdf:RDF (rdf:Description)> <![ATTLIST rdf:RDF xmlns rdf:CDATA #REQUIRED] <!ELEMENT rdf:Description (InstanceDescription)> <![ATTLIST rdf:Description about CDATA #REQUIRED] <!ELEMENT InstanceDescription (Management,Title,Description, Creator,Contributor, Subject, Publisher,Date, Identifier, Source,Relation, Coverage, Rights)> <!ELEMENT Management (RecordIdentifier, OriginatingSource, DRCreator)> <![ATTLIST InstanceDescription Type (Monographic Series Thesis Drawing Sketch Photograph) #REQUIRED Format (Text Image Audio Video) #REQUIRED Language (Chi Eng other) #REQUIRED] <!--ELEMENT Title (TitleProper, SubTitle?, Alternative?)*--> <!--ELEMENT Description (TableOfContents, Abstract, GeneralNotes, NotePertainingToPhysicalDescription)--> <!--ELEMENT Creator (NamePersonal NameCorporate NameConference)+--> <!--ELEMENT Contributor (CNamePersonal CNameCorporate CNameConference)+--> <!--ELEMENT Subject (STopic, STitle, SNamePersonal, SNameCorporate, SNameConference, SUncontrolledSubjectTerms, CLC?, DDC?)*--> <!--ELEMENT Publisher (Name, Place)--> <!--ELEMENT Date (Created, Issued)--> <!--ELEMENT Identifier (URI, URL, (ISBN ISSN)?, HoldingInformation?)*--> <!--ELEMENT Source (OriginalVersionInfo)--> <!--ELEMENT Relation (IsPartOf HasPart IsVersionOf HasVersionOf IsFormatOf HasFormatOf References IsReferencedBy IsBasedOn IsBasisFor Note SeriesTitle Requires)*--> <!--ELEMENT Coverage (Space, Time)--> <!--ELEMENT Rights (Copyright, IntellectualRight, Access, UserRightManagement)--> <!--ELEMENT RecordIdentifier (#PCDATA)--> <!--ELEMENT OriginatingSource (#PCDATA)--> <!--ELEMENT DRCreator (DRMaking, DRClean, ContentMark, MarkDirector, DataAudit, QC)--> </pre>	<pre> <!--ELEMENT DRMaking (#PCDATA)--> <!--ELEMENT DRClean (#PCDATA)--> <!--ELEMENT ContentMark (#PCDATA)--> <!--ELEMENT MarkDirector (#PCDATA)--> <!--ELEMENT DataAudit (#PCDATA)--> <!--ELEMENT QC (#PCDATA)--> <!--ELEMENT TitleProper (#PCDATA)--> <!--ELEMENT SubTitle (#PCDATA)--> <!--ELEMENT Alternative (#PCDATA)--> <!--ELEMENT TableOfContents (#PCDATA)--> <!--ELEMENT Abstract (#PCDATA)--> <!--ELEMENT GeneralNotes (#PCDATA)--> <!--ELEMENT NotePertainingToPhysicalDescription (Monographic, Drawing, Photograph)--> <!--ELEMENT Monographic EMPTY--> <!--ELEMENT Monographic MonPageNumber CDATA #REQUIRED MonHeight CDATA #REQUIRED--> <!--ELEMENT Drawing EMPTY--> <!--ELEMENT Drawing DrawingWidth CDATA #REQUIRED DrawingHeight CDATA #REQUIRED DrawingScale CDATA #REQUIRED--> <!--ELEMENT Photograph EMPTY--> <!--ELEMENT Photograph Width CDATA #REQUIRED Height CDATA #REQUIRED--> <!--ELEMENT NamePersonal (#PCDATA)--> <!--ELEMENT NameCorporate (#PCDATA)--> <!--ELEMENT NameConference (#PCDATA)--> <!--ELEMENT CNameCorporate (#PCDATA)--> <!--ELEMENT CNameCorporate (#PCDATA)--> <!--ELEMENT STopic (#PCDATA)--> <!--ELEMENT STitle (#PCDATA)--> <!--ELEMENT SNamePersonal (#PCDATA)--> <!--ELEMENT SNameCorporate (#PCDATA)--> <!--ELEMENT SNameConference (#PCDATA)--> </pre>	<pre> <!--ELEMENT SGeographic (#PCDATA)--> <!--ELEMENT SUncontrolledSubjectTerms (#PCDATA)--> <!--ELEMENT CLC (#PCDATA)--> <!--ELEMENT DDC (#PCDATA)--> <!--ELEMENT Name (#PCDATA)--> <!--ELEMENT Place (#PCDATA)--> <!--ELEMENT Created (#PCDATA)--> <!--ELEMENT Issued (#PCDATA)--> <!--ELEMENT URI (#PCDATA)--> <!--ELEMENT ISBN (#PCDATA)--> <!--ELEMENT ISSN (#PCDATA)--> <!--ELEMENT HoldingInformation (#PCDATA)--> <!--ELEMENT OriginalVersionInfo (#PCDATA)--> <!--ELEMENT IsPartOf (#PCDATA)--> <!--ELEMENT HasPart (#PCDATA)--> <!--ELEMENT IsVersionOf (#PCDATA)--> <!--ELEMENT HasVersionOf (#PCDATA)--> <!--ELEMENT IsFormatOf (#PCDATA)--> <!--ELEMENT References (#PCDATA)--> <!--ELEMENT IsReferencedBy (#PCDATA)--> <!--ELEMENT IsBasedOn (#PCDATA)--> <!--ELEMENT IsBasisFor (#PCDATA)--> <!--ELEMENT Note (#PCDATA)--> <!--ELEMENT SeriesTitle (#PCDATA)--> <!--ELEMENT Requires EMPTY--> <!--ELEMENT Requires GetStyle CDATA #IMPLIED SysRequire CDATA #IMPLIED--> <!--ELEMENT Space (#PCDATA)--> <!--ELEMENT Time (#PCDATA)--> <!--ELEMENT Copyright (#PCDATA)--> <!--ELEMENT IntellectualRight (#PCDATA)--> <!--ELEMENT Access (#PCDATA)--> <!--ELEMENT UserRightManagement (#PCDATA)--> </pre>
---	--	--

展为16个元素,并细化为66个子元素,建立了不同资源类型的 DTD 模板。其主要描述元素为:资源名称/Title,描述/Description,主要责任者/Creator,其他责任者/Contributors,主题/Subject,出版者/Publisher,日期/Date,资源类型/Type,资料格式/Format,资源唯一性标识/Identifier,资料出处/Source,语言/Language,关联/Relation,时空覆盖范畴/Coverage,数字式资源权限管理/Right,并在此基础上扩展了一项:系统管理/Management。我们制定了 THADL 元数据的中英文 DTD,本文只给出英文版,如详细表1。

3.1.3 Tamino 模式与 THADL DTD1.1 关系数据库中模式的定义是很重要的,同样 Tamino 定义了一组 DTD 来实现其模式(Schema)语言。该 DTD 包含了 Tamino 模式的三个元素:Collection,Doctype 和 Node。结构如表2-a, b 所示。

表2-a Tamino 模式 DTD

(!ELEMENT Collection (Doctype*))
(!ELEMENT Doctype (Node+))
(!ELEMENT Node EMPTY)

表2-b Tamino 和关系数据库对比

元素	类比	举例
Collection	数据库	医院
Doctype	表或视图	病人
Node	字段	姓名

(1) Tamino 模式的作用 Tamino 模式在存储时会记录一个规则,按照这个规则, Tamino 进行索引和存储数据;在查询时, Tamino 模式也提供对象合成器需要的相关信息去查询数据以满足用户提出的请求。这种 Tamino 模式也可以用图形编辑工具: Tamino 模式编辑器导入 DTD 后直接转换为 Tamino 的模式,转换的 Tamino 模式一部分值是来自 DTD,一部分是来自 Tamino 模式的默认值。同时可以针对不同的查询要求,对转换后的 Tamino 模式进行修改,以达到性能的改变。

(2) 节点的基本属性 在 Tamino 的模式中,最重要的是节点的属性,节点的基本属性: Name, Key, Parent。每个节点都有这三个基本的属性, Name 是这个节点的名称, Key 是一个唯一的识别码,它在整个数据库的模式中是唯一的, Parent 存放的是这一个节点的上一层节点的 Key 值,因此,是用来串联整个树状结构的关键。

节点的另外几个重要的属性值:表3是节点的 object-type 属性,表4是 Map-type 属性,表5是 Search-Type 属性,表6是 Map-Type="InfoField"时, Data-Type 相关的主要值,由于优化了存储,因此比 Map-Type="Native"效率高。

表3

Object-Type(节点类型常用的值有)	描述
CDATA	属性节点
PCDATA	带字符数据的元素节点
ANY	带任何内容的元素节点
CHOICE	带子元素的元素节点,并按照元素类型可以进行选择
EMPTY	没有内容的元素节点
SEQ	带有子元素的元素节点,有固定的顺序。一般是中间节点

表4

Map-Type	(用来说明节点的值以上那么样的类型存入 Tamino,常用的值有)
Native	以文字的方式存入
No	不存储
InfoField	存储,并且做索引
NonXML	这个节点的资料被当作非 XML 的资料来储存(即图像,影像,声音等)
Object	这个节点是映射到另外一个 XML Object,是用来做不同的 XML 文件之间的连接的

表5

Search-Type	(用来说明这个节点的索引方式,常用的值有)
text	文字索引,是叶子节点默认的索引方式
standard	标准索引,通常用于数值资料的索引
No	这个节点不作索引,通常用于在中间节点默认的索引方式
Standard text	文字索引与标准索引都做

表6

Data-Type	(用来说明这个节点的数据类型,常用的值有)
WCHAR	多字符的字符串,可变长度
FLOAT	单精度浮点值
INTEGER	整数(31位精度)

(3) THADL 元数据 DTD 1.1 与 Tamino 模式的转换 基于 Tamino 模式编辑器,我们开发了 THADL 元数据 DTD 到 Tamino 模式的转换模块,通过模块转换进行各种设置的修改,提供高性能的存储和查询服务。

3.2 THADL 元数据的实例

下面以蓟县独乐寺观音阁相关的照片、图纸和测稿为例(见图2-6),表7针对图2给出了基于 XML 和 RDF^[6]的元数据描述实例。



图2 观音阁南面照片



图3 观音阁南面对比照片



图4 观音阁南面透视图



图5 上层斗拱照片



图6 斗拱对比照片

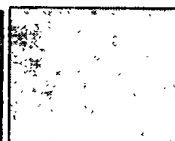


图7 上层斗拱测稿



图8 上层斗拱图纸

4 基于 Tamino 的元数据管理系统实现

我们已经开发了一个基于 Tamino 2.3.1.4 的 THADL 元数据管理的实现原型。前台客户端使用浏览器进行管理和服务, Web 服务器是 Apache 1.3, 后台在 Windows 2000 操作系统上采用 Tamino 数据库服务器, 基于 JSP^[9]和 Tamino client API-DOM API for Java, 构建了 THADL 元数据管理原型系统。系统管理和查询界面见图9~10, 主要实现的功能包括: 基于 XML 的查询和检索, 元数据的创建、添加、删除、修改等操作。

结束语 本文基于 XML 和 Tamino 数据库设计和开发清华大学建筑数字图书馆原型系统, 提出了一个基于 Tamino

的通用数字图书馆体系结构,设计了基于 XML 的 THADL 元数据 DTD,介绍了 THADL 原型试验系统的功能实现。原型试验系统的运行结果表明在存储和管理 XML 文档时,该

系统具有较高的效率和性能,同时运行稳定可靠,便于扩展。在构建 THADL 试验平台中我们也得到了以下体会。

表 7 蓟县独乐寺观音阁照片元数据描述实例

<pre><?xml version="1.0" encoding="gb2312" ?> <metadata xmlns:rdf="http://www.w3.org/1999/02/22- rdf-syntax-ns#"> <InstanceDescription Format=" 图片" Language="cn" Type=" 照片"> <rdf:RDF> <rdf:Description about="07-yzxshk-03-165-f3.jpg"> <Management> <RecordIdentifier /> <OriginatingSource /> <DRCreator> <DRMaking>清华大学建筑数字图书馆项目组 <DRMaking> <DRClean>林洙<DRClean> <ContentMark>罗德胤</ContentMark> <MarkDirector>郭黛姮</MarkDirector> <DataAudit>周虹</DataAudit> <QC>吴开华</QC> <DRCreator> <Management> <Title> <TitleProper>独乐寺观音阁南面</TitleProper> <SubTitle /> <Alternative /> <Title> <Description> <TableOfContents /> <Abstract /> <GeneralNotes /> <NotePertainingToPhysicalDescription> <Monographic MonHeight="" MonPageNumber="" /> <Drawing DrawingHeight="" DrawingScale="" DrawingWidth="" /></pre>	<pre><Photograph Height="" Width="" /> </NotePertainingToPhysicalDescription> </Description> <Creator> <NamePersonal>梁思成</NamePersonal> <NameCorporate> 中国营造学社 </NameCorporate> </NameConference /> </Creator> <Contributor> <NamePersonal /> <NameCorporate /> <NameConference /> </Contributor /> <Subject> <STopic>独乐寺;观音阁;寺院;南面;大 屋顶;楼阁式建筑;歇山;平坐;照片 </STopic> <STitle>营造学社汇刊;蓟县独乐寺观音阁山 门考</STitle> <SNamePersonal>梁思成</SNamePersonal> <SNameCorporate> 中国营造学社 </SNameCorporate> <SNameConference /> <SGeographic>蓟县</SGeographic> <SUncontrolledSubject Terms /> <CLC>TU-098.321</CLC> <DDC /> </Subject> <Publisher> <Name>清华大学营造学社数字图书馆项目 小组</Name> <Place>中国北京</Place> </Publisher></pre>	<pre><Date> <Created>2000.12.07</Created> </Issued /> </Date> <Identifier> <URI /> <URL>07-yzxshk-03-165- f3.jpg</URI> </ISBN /> </ISSN /> <HoldingInformation /> </Identifier> <Source> <OriginalVersionInfo /> </Source> <Relation> <IsPartOf /> </Relation> <Coverage> <Space>蓟县</Space> <Time>建于辽圣宗统和 . 年 (984)</Time> </Coverage> <Rights> <Copyright> 中国营造学社 </Copyright> <IntellectualRight /> <Access /> <UserRightManagement>1</UserRi ghtManagement> </Rights> </rdf:Description> </rdf:RDF> </InstanceDescription> </metadata></pre>
--	--	---

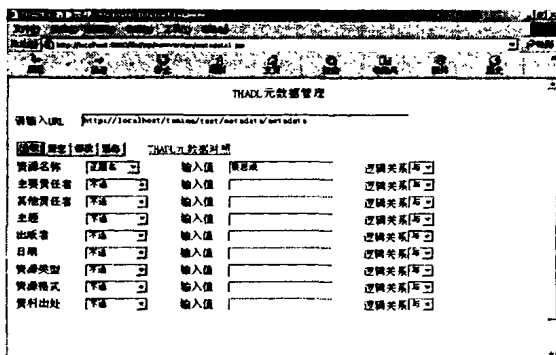


图 9 THADL 的查询界面

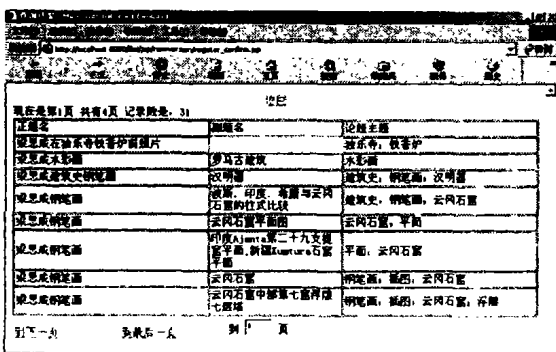


图 10 查询结果界面

(1) 建立基于 XML 数据库的应用中,基于 DTD 或者 Schema 建立 Tamino 模式是很关键的一步,因为它确定了 Tamino 内部存储和查询的规则,在此基础上才能更好地建立应用,才能使查询更加迅速。

(2) 由于 Tamino 的查询语言 X-Query 的查询的方法很

多,因此如何查询才能优化,也是将来要进一步深入研究的问题。

(3) 采用 Tamino Client API-DOM API for Java,由于 DOM 处理文档时,把整个文档放进内存,会很占用内存。将来应该结合 DOM 和 SAX 一起开发应用程序。

(4) 未来将与德国 Software AG 公司合作,开展支持 XML 关系数据库(如 ORACLE 9i)和 Tamino 新版的大量测试工作,研究和开发功能更强、性能更高的 XML 数据库管理系统。

参考文献

- Xing Chunxiao, Zhou Lizhu, et al. THADL: A Digital Library for Chinese Architecture Study. In: The 12th Intl. Conf. on New Information Technology, Beijing, China, 2001. 373~382
- Yang C C, Chan W W M. Metadata Design for Chinese Medicine Digital Library Using XML. In: Proc. of the 33rd Hawaii Intl. Conf. on System Sciences, 2000. 1~10
- Tidwell D. XML Programming in Java Cyber Evangelist, Developer Works XML Team. 1999
- Tamino Information Server White Paper and Tamino Architecture. <http://www.softwareg.com/tamino>
- The State of the Dublin Core Metadata Initiative. D-Lib Magazine, Vol. 5 No4, April 1999. <http://www.dlib.org/dlib/april99/04weibel.html>
- Resource Description Framework (RDF). <http://www.w3.org/RDF/>
- Bourret R. XML Database Products. 2000. <http://www.rpbourret.com/xml/XMLDatabaseProds.htm>
- Elliotte Rusty Harold. XML 实用大全. 中国水利水电出版社, 2000
- 黄浩文, 黄静舒. JSP 核心技术和电子商务应用实例. 机械工业出版社, 2001