

基于地理本体的不可移动文物信息参考模型研究及应用^{*}

吴琼^{1,2} 鲍泓²

(北京交通大学计算机与信息技术学院 北京 100044)¹ (北京联合大学信息技术研究所 北京 100101)²

摘要 数字博物馆是建立以存储、管理和展示文物为目的的信息系统,以往的大都是面向可移动文物的。本文研究基于地理本体的数据表示模型,基于本体的方法可以为不可移动文物添加丰富的语义信息,而 GIS 可以为不可移动文物提供地理信息上的支持,结合两者,将可以构建面向不可移动文物的信息参考模型。

关键词 CIDOC CRM, 地理本体, 融合, 不可移动文物

Investigation and Application of Geospatial Ontology-based Information Model for Immovable Cultural Relic

WU Qiong^{1,2} BAO Hong²

(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044)¹

(Institute of Information Technology, Beijing Union University, Beijing 100101)²

Abstract In the past, the digital museums were mostly information systems for the purpose of management of movable cultural relics. Ontology-based data representation model can provide abundant semantic information for immovable cultural relic, and GIS can provide geospatial information for it. With combining ontology-based data representation model and GIS, we can establish the ontology-based digital museum of immovable cultural relic.

Keywords CIDOC CRM, Geospatial ontology, Combination, Immovable cultural relic

1 引言

文物是人类社会活动中遗留下来的具有历史、艺术、科学价值的遗物和遗迹,可以分为可移动文物和不可移动文物。我国的历史文化遗产众多,利用现代计算机技术实现文物的数字化,有利于文物的保护、管理和利用。随着城市建设改造和国家大型基础资源项目的开发建设,如何有效地保护、管理和利用不可移动文物,是我们迫切需要解决的问题。由于不可移动文物作为地理实体还包含着与其相对应的地理信息,因此我们通过将 GIS 与不可移动文物的本体模型相结合,建立了一个面向不可移动文物的数字博物馆,可以较好地解决上述问题。

2 本体及 GIS 简介

2.1 本体

本体(ontology)是一个哲学上的概念,是指哲学中研究世界的本原或本性的部分。从本体论的基本概念出发,哲学上把本体定义为“对世界上客观事物所进行的系统描述”。本体在计算机领域是指定义元数据及其相关关系的规范。它依赖于所采用的语言,按照表示和描述的形式化程度不同,可以分为完全非形式化、半形式化和严格形式化。本体的形式化程度越高,越有利于计算机进行自动处理^[1]。

本体最为突出的应用是智能信息检索。基于本体的智能检索信息系统由于能提供查询和资源描述所必需的元语,并通过领域语义模型为信息源提供语义标注信息,从而使系统内所有的 Agent 对领域内的概念、概念之间的联系及领域内的基本公理知识有一个统一的认识,进一步提高了系统的联

想能力和精确性,为用户提供有价值的信息和一个全面的共同视图。

本体在信息检索领域的另一个应用就是语义网的开发。在语义网中,所有信息都有明确的语义信息,从而自动对网络上的资源进行处理,使网络信息得到更充分的利用。本体具有良好的概念层次结构并且支持知识的推理,因此本体成为构建语义网的核心内容。

在数字博物馆领域引入本体的直接目的是解决数据标准规范的定义与数据描述语言规范化的问题。国际上对博物馆领域本体的研究已经有一段较长的时间,尽管没有一个统一的标准,但是研究已经取得了很多的成果。目前,国际上一些团体或相关组织已经制定了文化遗产领域的本体模型,包括:CIDOC CRM Conceptual Reference Model (CRM)模型^[2]、ABC ONTOLOGY^[3]模型以及 CDWA 模型^[4]等。其中 CIDOC CRM 是这些标准中较为全面、严格遵循本体原则的一个模型,它有超过 10 年的历史,被世界上多个很有影响的项目所采用。作者所在的语义网项目研究小组已经在此方面做了一些研究工作^[5~8]。

2.2 GIS 及地理本体

地理信息系统(Geographical Information System,简称 GIS)一般定义为基于计算机方法的空间信息获取、组织和分析技术,它具有数据库管理、制图、图像处理和分析功能。GIS 使用空间的、描述性的数据通过地图来定义物体及其真实世界的特征,如建筑物、河流或道路,能提供多维空间的实际效果。表现数据的空间特征是 GIS 明显区别于其它信息系统或传统的数据库和电子数据表的一个明显特征。

空间特征数据记录的是空间实体的位置、拓扑关系和几

^{*}北京市科委北京市科技计划(Z0006317041691)、北京市教育委员会科技发展计划项目(km200611417001)。吴琼 研究生;鲍泓 教授,主要从事分布式系统、软件工程研究。

何特征。空间特征指空间物体的位置、形状和大小等几何特征,以及与相邻物体的拓扑关系。物体的空间特征信息也就是它的地理信息。空间位置可以由不同的坐标系统,如经纬度坐标、一些标准的地图投影坐标或是任意的直角坐标等来描述。对空间目标的定位一般不是通过记录其空间坐标,而是确定某一目标与其它更熟悉或已知的目标间的空间位置关系,而这种关系往往也是拓扑关系。如某一遗址位于某一河流、山脉或道路的某一位置。不可移动的文物数据都具有空间特征,表现不可移动文物的空间特征正是 GIS 技术的优势所在。

地理本体就是将有关地理科学领域的知识、信息和数据抽象成一个个具有共识的对象(如实体),并按照一定的关系组成体系,同时进行概念化处理和明确定义,最后以形式化表达的理论和方法^[9]。地理信息的本体论研究在上个世纪 90

年代中后期逐步得到重视。构造地理本体的主要目的是捕获相关地理领域的知识,提供对该领域知识的共同理解,确定该领域内共同认可的词汇,并从不同层次的形式化模式上给出这些地理词汇和词汇间相互关系的明确定义,从而为该领域提供语义和知识层次上描述地理信息系统的概念模型建模工具。目前,国内外已有不少研究机构和学者对地理信息科学中的本体论研究极为关注,做了一些研究工作^[9~13]。但是在地理本体的研究与构建上,无论是国内还是国外,许多关键问题都没有达成共识,也没有一个成熟的规范。因此,在我们的实际应用中,有选择性地构建了符合应用需要的地理本体,完成对不可移动文物地理信息的描述。

3 基于 CIDOC CRM 和地理本体的信息参考模型

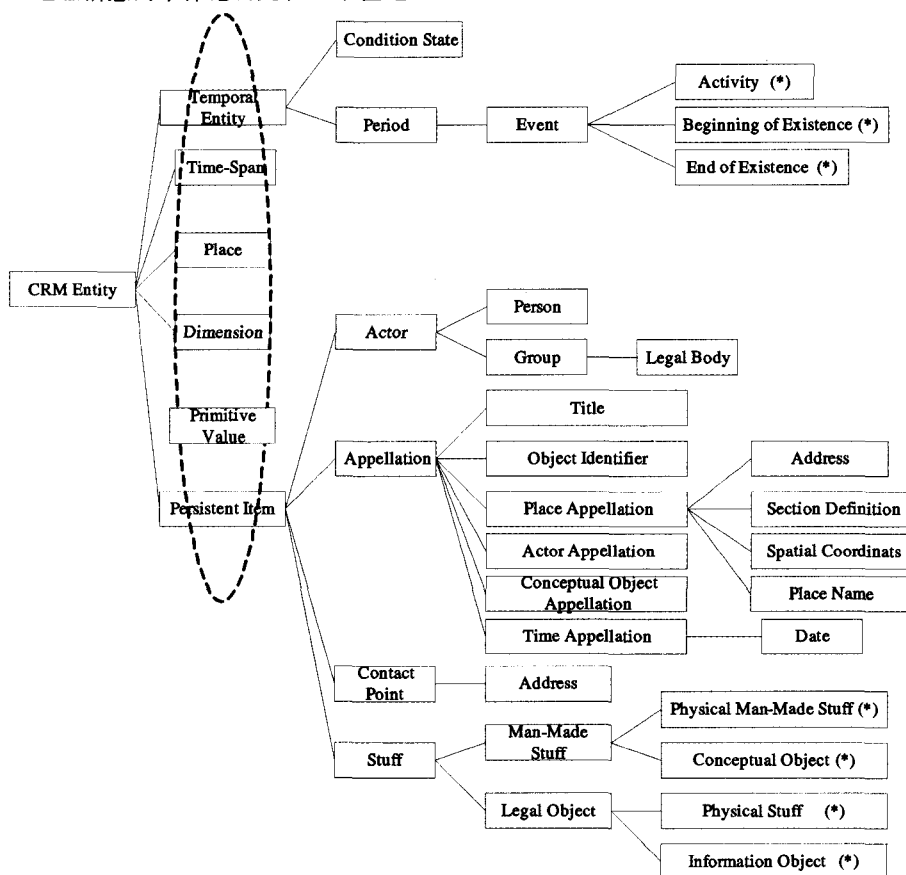


图 1 CIDOC CRM 部分实体及实体间继承关系

3.1 CIDOC CRM 模型

CIDOC CRM 是 CIDOC 组成 ICOM(International Council for Museums)组织的 20 几个委员会(committee)之一,全名是 Comité International pour la DOCumentation, 简称为 CI-DOC。它是一个可扩展的领域本体模型。CIDOC CRM 模型涵盖了围绕特定文物发生的历史史实、人文艺术、考古遗迹、时间、地点、人物以及版权声明等信息,由于有较长的历史,因此发展得比较完善。

最新的 CIDOC CRM V3.4.9 版本将文化遗产领域定义出 84 个实体(entity)和 141 个属性(property)^[2]。而且,随着模型的不断扩展,实体和属性数量可能还会增加。整个 CI-DOC CRM 架构以树的结构呈现,部分实体具有多重继承的特性。每一个实体都具有属性,而实体除了拥有自己的属性

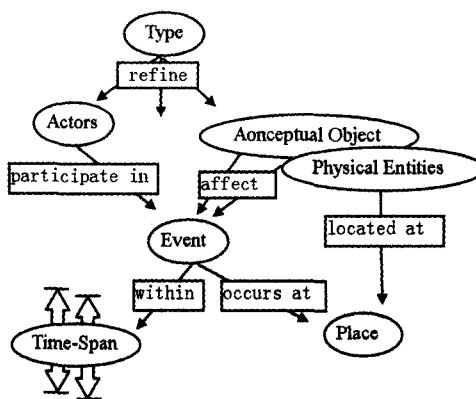


图 2 CIDOC CRM 实体与属性关系

之外,还会继承其父节点的所有属性。实体通过它的属性连接到另外的实体,以实现对某个特定文物的描述。图 1 中所列出的是 CIDOC CRM 的部分实体以及实体间继承关系。

从图 2 我们可以看出, CIDOC CRM 是如何描述文化遗产领域中的实体以及它们之间的相互关系的。例如: Physical Entities 实体和 Place 实体有着“located at”的关系, Physical Entities 实体和 Event 实体有着“affect”的关系, Actor 实体又和 Event 实体有着“participate in”的关系。利用这个架构,

我们可能有的描述有:

- [故宫]属于[Physical Entities]
- [北京]属于[Place]
- [故宫修建工程]属于[Event]
- [中国工匠]属于[Actors]
- [故宫]located at [北京]
- [故宫]affect [故宫修建工程]
- [中国工匠] participate in [故宫修建工程]

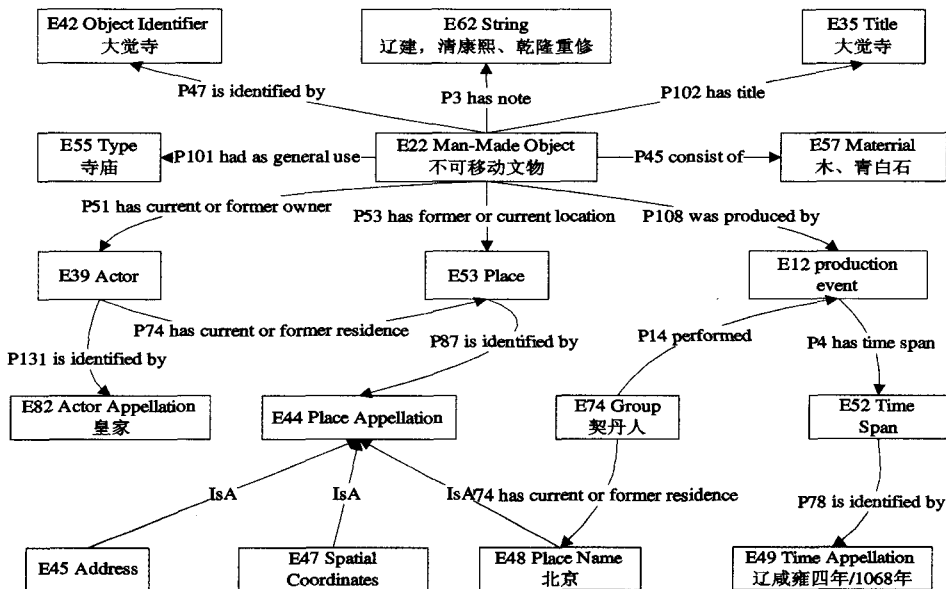


图 3 CIDOC CRM 对大觉寺的描述

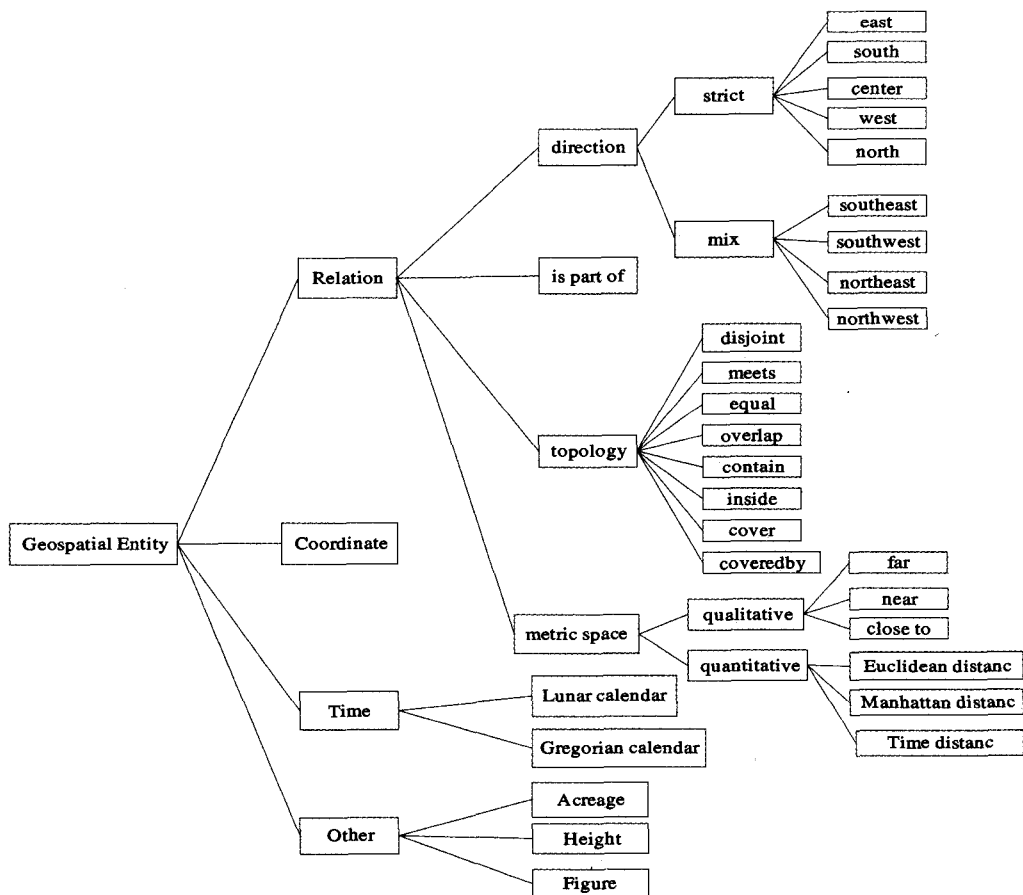


图 4 地理本体模型

以北京市重点文物保护单位大觉寺为例,应用 CIDOC CRM 模型对大觉寺的语义信息描述如图 3 所示。该模型中所有的实体和属性均取自 CIDOC CRM 模型。由于实体间通过属性关系连接起来,因此通过这种方式,不但可以表示数据,还可以表示数据之间的联系,体现了丰富的语义信息。CIDOC CRM 是文化遗产领域的整体规范,而我们的实际系统不一定涵盖所有它所描述的信息,所以我们根据实际需求选择了部分实体和属性关系。

3.2 地理本体模型

对于一个地理实体来说,它的地理信息主要体现在它的空间位置以及与其他地理实体的关系。空间位置一般有两种方法定义:一种是按照大地参照系坐标定义,如经纬度坐标;一种是按照相对位置关系定义,如空间上的相邻、包含等关系。地理实体之间的关系主要有拓扑关系、顺序空间关系、度量空间关系、整体-部分关系等。因此对于一个地理实体,可以建立一个如图 4 所示的地理本体。

应用以上地理本体模型,可以对大觉寺的地理信息进行描述,如图 5。根据实际情况,我们也是选取了部分实体和属性。

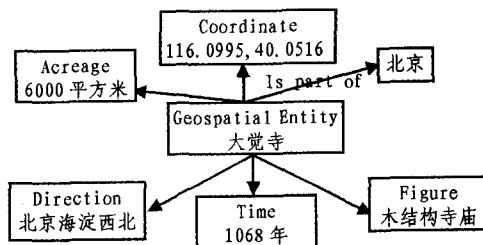


图 5 地理本体对大觉寺的描述

3.3 基于 CIDOC CRM 和地理本体的信息参考模型的应用

通过前面的示例我们可以看出,CIDOC CRM 模型可以很好地描述围绕不可移动文物发生的历史史实、时间、地点、人物等语义信息,但是它的整体模型不支持对不可移动文物的位置、形状和大小,以及与相邻地理实体的拓扑关系这些地理信息的描述。同样,地理本体也无法对不可移动文物的语义信息进行描述。因此,我们需要将这两者融合,从而构建针对不可移动文物的元数据模型,以满足实际需求。

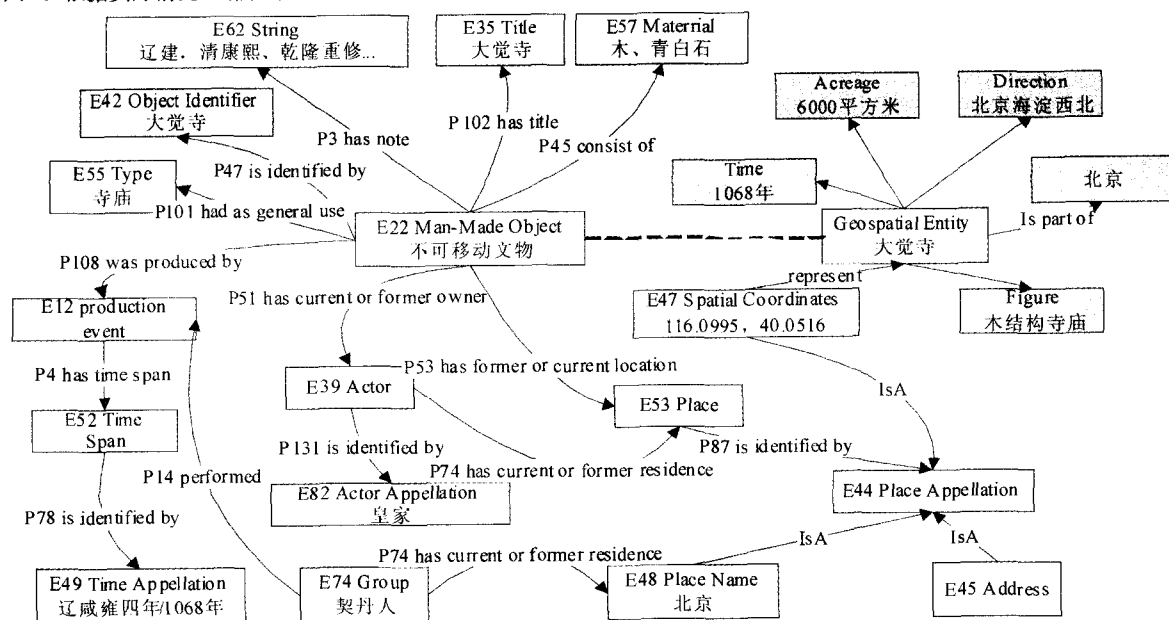


图 6 基于 CIDOC CRM 和地理本体对大觉寺的描述

如图 6 所示,我们通过 CIDOC CRM 的实体 E47 Spatial Coordinate 和地理本体进行连接,对 CIDOC CRM 进行扩充。由于在前面的地理本体中,坐标项 Coordinates 也用于描述不可移动文物的坐标,因此这使得连接成为可能(浅蓝色部分为 CIDOC CRM 模型,浅黄色部分为地理本体模型)。可以看出,这种元数据模型不仅能够提供文物本身的语义特征信息(例如:大觉寺的修建时间、位置、修建大觉寺的群体等),还能够提供文物的地理信息(例如:大觉寺的占地面积、所属城市及所在方位、坐标等),使得使用统一的元数据模型描述不可移动文物的信息成为可能。

4 基于地理本体的系统构建方法

4.1 总体架构设计

系统如图 7,主要包括以下主要组成部分:

- 图形化用户界面;
- 不可移动文物本体模型;
- 命令分发模块;

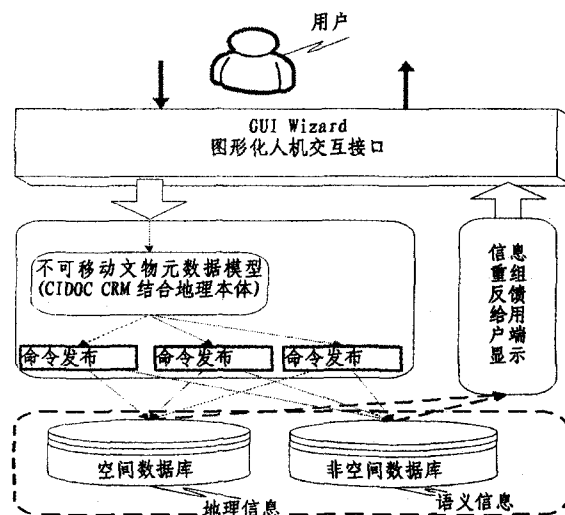


图 7 系统整体架构

的操作,从而减少了时间和空间消耗。ARSDM 算法以约简条件为启发信息,采用边得到矩阵元素边删除 ΔA 中元素边排序的方法,有效地节省了约简的时间和空间消耗,提高了约简的效率。从实验结果可知,ARSDM 算法对较大规模数据库的处理比经典算法有明显优势。

参 考 文 献

- 1 Pawlak Z. Rough Sets [J]. International Journal of Computer and Information Science, 1982, 11(5): 341~356
- 2 Pawlak Z, et al. Rough Set [J]. Communication of the ACM, 1995, 38(11): 89~95
- 3 徐捷,徐从富,耿卫东. 基于粗糙集理论的动态目标识别及跟踪 [J]. 电子学报, 2002, 30(4): 605~607

- 4 赵斌,黄翠微,叶东毅等. 粗糙集中属性约简的一个贪心算法 [J]. 系统工程与电子技术, 2000, 22(9): 63~65
- 5 王国胤,于洪,杨大春. 基于条件信息熵的决策表约简 [J]. 计算机学报, 2002, 25(7): 759~766
- 6 王清毅,范焱,蔡庆生. 知识的约简研究 [J]. 小型微型计算机系统, 2000, 21(6): 623~627
- 7 谭天乐,宋执环,李平. 信息系统数据清洗. 规则提取的矩阵算法 [J]. 信息与控制, 2003, 32(4): 289~294
- 8 Skowron A, Rauszer C. The discernibility matrices and functions in information system [A]. Intelligent Decision Support Handbook of Applications and Advances of the Rough Sets Theory [C], Dordrecht: Kluwer Academic Publishers, 1992. 331~338
- 9 <http://www.ics.uci.edu/~mlearn/databases/>

(上接第 116 页)

· 数据源: 地理信息数据库, 语义信息数据库。

4.2 系统的主要功能

4.2.1 映射与命令分发

当用户发出请求后,请求首先到达系统内置的元数据模型(CIDOC CRM 与地理本体结合实现)。通过元数据模型到各个数据库的映射机制,对请求进行转换,分解成针对各个数据库的命令,发送到数据库层。对数据库进行操作后,所得到的信息再反馈回用户端,显示给用户。

4.2.2 信息浏览

通过图形化用户界面,系统实现与用户的交互。用户可以在 GIS 数据库提供的地图上直观地看到各个文物点的分布,并且选择浏览他感兴趣的文物的相关信息。与文物相关的语义信息由语义信息数据库提供。

4.2.3 信息查询

提供各类数据的查询功能,如提供文物保护相关数据的查询,包括数据的简单查询(根据文物类别,选择要查询的数据表和字段,获取用户需要的数据);数据综合查询(通过指定数据查询类型,任意组合数据过滤条件,获取用户所需要的数据)。

4.2.4 信息发布功能

可通过互联网工具进行一些公众文物保护信息和相关法律、法规信息的发布等。

结论 本文从文物领域本体模型 CIDOC CRM 对不可移动文物数据描述的局限性出发,分析了 CIDOC CRM 和地理本体两个元数据模型对不可移动文物的描述方式,并提出了将两者融合的方法,给出了实例并加以说明。同时,结合了 GIS 在表现不可移动文物空间特征的优势,提出了基于地理本体的不可移动文物数字博物馆的系统架构,为不可移动文物的保护管理提供了新的思路 and 手段。

参 考 文 献

- 1 William S, Austin T. Ontologies. IEEE Intelligent Systems, Jan./Feb. 1999. 18~19
- 2 CIDOC Conceptual Reference Model. <http://cidoc.ics.forth.gr/>
- 3 Brickley D, Hunter J, Lagoze C. ABC: A Logical Model for Metadata Interoperability? Harmony Project, Working Paper. 1999. http://www.ilt.bris.ac.uk/discovery/harmony/docs/abc/abc_draft.html
- 4 The Getty Information Institute and College Art Association. Categories for the Description of Works of Art(CDWA). <http://www.getty.edu/gri/standard/cdwa/>
- 5 Liu Hong-zhe, Bao Hong. An Information Flow-based Approach to Semantic Integration of Distributed Digital Museums. In: Proceedings Machine Learning and Cybernetics, Volume 7 of 7, 2006. 4430~4437
- 6 Bao Hong, Liu Hongzhe, Yu Jiehua. Ontology-based Semantic Integration for Digital Museums. In: Lecture Notes in Computer Science (LNCS), 3739 Chapter, Springer, 2005. 0302~9743
- 7 刘宏哲,鲍泓,余杰华. 基于 CIDOC CRM 的虚拟博物馆语义网络架构. 计算机应用研究, 2006(3): 050~054
- 8 鲍泓,刘宏哲. 基于 Web Services 的虚拟文物博物馆架构. 系统仿真学报, 2005, 17(6): 1412~1417
- 9 陈建军,周成虎,王敬贵. 地理本体的研究进展与分析. 地学前沿, 2006, 13(3): 081~090
- 10 Wang Jingui, Fen Zhen. Oceanographic ontology-based spatial knowledge query. Acta Oceanologica Sinica, 2005, 24(4): 66~71
- 11 Fonseca F. Ontology-driven Geographic Information Systems. The University of Maine, 2001
- 12 Fonseca F, Egenhofer M, Agouris P, et al. Using ontologies for integrated geographic information systems. Transactions in GIS, 2002, 6(3): 231~257