

地理属性和图形属性耦合的信息编码模型

焦 健 曾琪明

(北京大学遥感与 GIS 研究所 北京 100871)

摘 要 对地理信息进行编码,是 GIS 设计中最重要的一步,它代表着现实世界与信息世界的界面。本文从地图的模型化和空间认知过程出发,提出地理实体编码模型与图形实体编码模型的概念,并着重分析了代表这两类不同编码方式的软件对制图功能的影响,并在此基础上提出了地理属性和地图图形属性耦合的信息编码概念模型。

关键词 信息编码模型,地理信息系统,地图制图

目前尽管对地理信息系统(GIS)的定义有着众多的分歧^[1],但是 GIS 的内涵之一是空间信息的载体,对此人们基本达成共识。地图具有存储和传输空间信息的能力,也是空间信息的载体。GIS 脱胎于地图^[2],二者在许多方面具有相同的内涵和很深的渊源。地图设计与制作功能是衡量与评价 GIS 的重要指标之一,因而无论是 GIS 的开发者还是用户,都非常重视 GIS 生产高质量地图的能力。为了加强和完善 GIS 的制图功能,进行了多方面的研究,但大多是从功能的覆盖面、算法、硬件设备及配套的功能等方面进行。本文从信息编码的角度,分析了不同的信息编码方法对 GIS 制图功能的影响,进而提出了新的信息编码模型,以期增强 GIS 的制图功能。

1 地理属性信息编码模型

一般而言,地理信息主要包括 3 大类:反映空间实体位置的几何信息;描述实体性质和特征的属性信息;和反映空间实体在时间域内变化的时间信息。通常将属性信息分为定性与定量两种形式。定性信息用来描述要素的分类或对要素进行标名;定量信息反映了空间实体数量方面的特征。把定性与定量信息结合起来,进行分类和分等定级,依类依级编码所构成的模型称为地理属性信息编码模型。

从用户所认知的地理现象结构,到记录在计算机存储介质上的地理数据,一般有 4 个阶段^[3](据 P. A. Burrough 修改)(图 1)。

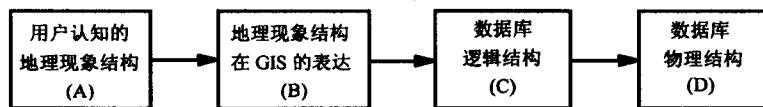


图1 描述地理信息的4个阶段

Fig.1 Four steps to describe geographical information

从(A)到(B),即对地理信息进行编码,是 GIS 设计中最重要的一步,它代表着现实世界与信息世界的界面^[4]。最典型的地理属性信息编码模型是按树型分类结构和多维指标结构构建的^[5]。一般分类系统大多属于树型结构的分类体系,例如土地利用分类系统、旅游资源分类系统等,所以一般 GIS 的属性数据大多也属于树型结构。树型分类结构是随着级别的层次不同,要用数位代码来表示一种空

间实体属性,例如,耕地在一级土地利用分类中,代码为 1,代码个数是 1。水稻田属于第 3 级分类,代码为 111,代码个数是 3;多维指标结构编码模型中的每一位代码,表示一种属性,例如,在土壤侵蚀分类中,可以用第 1 位码表示侵蚀动力类型,第 2 位码表示侵蚀强度,第 3 位表示抗侵蚀年限,……。这两种结构的数据模型,对 GIS 的数据采集、空间分析与运算等方面,均有不同的影响。相对而言,树型分类结

构的地图图斑类别确定,要比多维指标结构的图斑确定容易一些,因为后者一个图斑内可能包涵数个属性特征,借助于遥感图像的分析和解译,可拓展解决这个问题的途径。在GIS中,由多维指标结构的数据,很容易派生不同内容的专题地图。

虽然上述两种模型,对GIS的数据获取、存储、分析、制图上各具千秋,但它们都是仅基于地学概念进行属性信息编码的。

2 地图的模型化和空间认知过程对编码的启示

根据模型论原理,当一个系统与另一个系统,在某些性质方面可以建立起某种相似性,那么二者之间就存在着原型与模型的关系。其中一个被看作是系统的原型,另一个则为模型。对于一个系统,可以从许多不同的角度,用抽象概括的方法,将系统各方面的一些主要性质抽象出来构造模型,通过对模型的研究以取代对原型的研究。因此,对于一个系统,可以构造多个模型。对多个不同模型的研究,就可以获得对原型系统的相对完整的认识。

从模型论的观点定义地图:地图是空间要素和有关地表知识信息的有组织图像模型^[6](据C. 博德)。做为模型的地图,是制图人员把对地理环境(原型)的认知,通过选取,分类和简化,用地图语言来描述,形成模拟地图(图像模型)。把模拟地图输入到计算机中,成为计算机可接受的形式,它是一种数字地图(数字模型)。根据模型论原理,模拟地图和数字地图,都是同一空间实体不同形式模型。那么,根据图像模型构造新的数字模型应该也有不同的方法,上述从地理学角度研究属性编码仅为这类构模方法中的一种。为了探求新的编码方法,本文从认知科学的角度研究地图认知过程,以期得到启示。

制图者从一种复杂的未经组织的外部环境中选取、抽象和组织信息,把它们转变为有组织的知识结构,并在地图上用符号组合成相关图形。用图者通过识别不同的符号与符号组合,在其头脑中重构环境信息的空间关系,借助于内在的空间认知能力转化为用图者关于环境的认知。这是人们通过地图获得空间认知的一个完整过程。

阅读地图是实现空间认知最有效的手段^[7]。地图的空间认知过程如图2。

对于一个地理学家或有经验的地图读者,他可以通过地图符号迅速从地图上获得有关河流的类

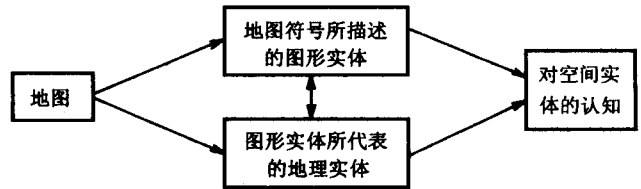


图2 地图的空间认知过程

Fig.2 Spatial recognition process from map

型、山脉的走向、城镇的布局等地理学概念。这里地图符号仅仅起了一个短暂的桥梁作用,读者的注意力主要集中在地图符号所代表的地理实体上。

对于一个空间认知能力较低或无读图经验的读者,首先吸引他注意力的是各种地图符号。他必须首先阅读图例,理解每种符号所代表的含义,然后再阅读符号的组合,分析图形结构,才能获得一定的地理知识。这里地图符号起着极为重要的桥梁作用。

上述地图的空间认知过程给我们启示:地图所传递的信息可以从二个层次上去理解,即地图符号所描述的图形实体层和图形实体所代表的地理实体层。地理实体层可以通过图形实体层反映出来,它只有当真实空间世界的实体发生变化时才改变,相对而言比较稳定。图形实体层随着地理实体层的变化而变化,也可由地图设计者的主观意志所决定。采用不同地图表示方法,采用不同的地图符号,改变地图图面配置格局等,都会使图形实体发生变化。不论是地图的地理实体层,还是图形实体层,都可认作是对同一真实世界从不同角度抽象的模型^[8]。

3 地理实体编码与图形实体编码模型

对于地理信息的编码,笔者认为也可从两方面进行。从地图的地理实体层的角度进行编码,称为地理实体编码模型(Geographical Entity Encoding Model-GEEM)。即用属性特征码、坐标及空间关系描述存贮的对象。上述的树型分类结构与多维指标结构编码模型都属于此类。二是从地图的图形实体层的角度进行编码,称为图形实体编码模型(Cartographic Entity Encoding Model-CEEM)。它是用地图符号的图形属性码、坐标和符号间的关系来描述地理对象,是地图生产者特有的模型。

根据 Jacques Bertin (1967年)的研究,任何图形都可分解为几个基本元素,称为视觉变量。A. H. Robinson 等人根据地图的特点将其归纳和扩充为7种基本图形要素:色相、亮度、尺寸、形状、密度、方向

和位置^[9]。这些基本图形要素的各种组合,就可构成千变万化的地图图形。因此,为了直观有效地表现地理数据,可通过改变符号的外观,系统地调节基本图形要素,使地图上的符号变得比较有特色和突出。这样有关 Bertin 的视觉变量或 Robinson 的基本图形要素理论,就可做为确定图形属性码位的理论基础。

依据地图符号在地图上的尺度及形状,它又可被分为点状(注记可归于此类)、线状和面状符号 3 种类型。分解每种形式的符号,研究其构图规律,确定图形属性的码位和含义,对这些码位赋予一定的参数,则可建立地图图形实体编码模型。例如,地图图斑是由一系列坐标值所决定的面状符号,其主要的构图要素是颜色和内部纹理。当符号内的颜色和纹理一经确定,该面状符号基本被确定。因此,面状符号的图形属性码可以包括:颜色代码、图元库代码、图元尺寸(mm)、图元行角度($^{\circ}$)、图元行间隔(mm)、图元列间隔(mm)、行列交错否(Y/N)。或者仅仅是颜色代码和填充图案。

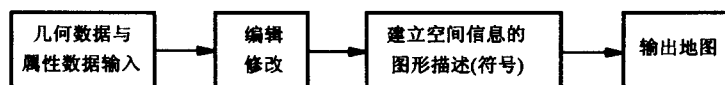


图3 ARC/INFO的制图模式

Fig.3 Mapping scheme by using ARC/INFO

先输入空间实体的位置和属性特征信息,进行编辑,修改错误,然后对属性信息用符号(形状、尺寸、颜色等)可视化表示。

ARC/INFO 的这种制图模式有如下优点:

易于派生新图和更新地图

ARC/INFO 用属性码、坐标与空间关系来描述存贮对象,是面向空间实体的。它不规定用什么符号系统来具体表示,因而它存储的数据是独立于制图表示方法的。这种信息的存贮与它们在介质上的可视符号表示分开的方法,提高了数据的检索与图形表示的灵活性。借助于符号库、颜色库、线型库等,可以随时抽取数据生成满足某种需要的地图。

编制新图首先需要通过各种空间数据分析与处理方法,提取和派生出新的信息;更新地图也首先需要更新图形库与属性库中的数据。ARC/INFO 这种描述存贮对象和对空间数据管理的方法,使得派生专题图和更新地图比较容易。

便于快速、多样地确定地图的表现形式

ARC/INFO 存贮的信息与其可视化符号分开,二者通过查找表(Look-Up Table)联系^[10]。查找表

4 二种编码方式的软件在地图制图上的特色

纵观国内外 GIS 和地图制图软件,其中大多数基本上可被认为是基于地理实体编码模型进行设计的,比较有代表性的是 ARC/INFO,是由美国环境系统研究所(ERSI)开发的一种通用 GIS 软件。只有极少数的制图软件可被认为采用的是图形实体编码模型,如 AutoCad、MapCad 等,后者是中国地质大学计算中心开发的应用于地学制图的专用软件。下面以 ARC/INFO 和 MapCad 为例,分析这二类软件因编码方式不同所形成的制图上的特色。

4.1 PC ARC/INFO 的制图特点

忽略其制作地图的具体操作与步骤,其基本过程可概括为如下模式(图 3)。

这也是一般 GIS 和某些地图制图系统(Mapping System—MS)制图的模式。它们的共同特点是:

是一种特殊的数据文件,它至少包含两项:一是与属性表中的某项相同,其值做为查找表中的关键字;另一项为符号项。使用查找表有如下优点:第 1 可以用任何适当的符号来反映属性特征,即对不同的属性采用不同的符号,从而快速确定了地图内容的表示方法。它很容易或者只改变符号代码,就可制作出不同风格和特色的地图;第 2,利用查找表可以减少数据的存贮量。在查找表中,同一类的属性值只存贮一次符号码,而不是对属性表中的每一个记录都赋予符号代码。查找表的作用如同地图中的图例,它是整个制图区域所有符号的归纳与说明,图例的符号决定了整个制图区域内的同类属性的符号。

ARC/INFO 具有强大的制图功能,但有时仍感到不便与繁琐。“所见即所得”是地图制图系统设计者追求的目标。它不仅仅体现在颜色上即设计人员设色—屏幕颜色—输出地图颜色的一致性上,而且应该包括符号显示上的“所见即所得”。制图者面对计算机屏幕应该像面对模拟地图一样,屏幕显示的地图可以是最终输出地图的式样。只有这样,制图者才能集中精力致力于提高地图的科学性与艺术性

上。但是,ARC/INFO对图形数据的编辑和符号设计过程,分离成编辑与制图两种独立状态。这里编辑状态是指为了进行编辑修改,地图要素没有以设计的输出符号表示的一种状态。此时屏幕上显示的点状或线状符号等,实际上仅反映了制图要素的空间位置、形状和关系,其属性要素没有得到直接的图解反映;制图状态是指属性数据以直观的、最终输出地图时的图解形式表现出来,即用不同形状的符号、线型、颜色、图案等表现全部的制图要素,是地图输出前的状态。地图的基本特性之一就是采用符号表示而具有直观性。不同特征的空间要素用不同的符号、颜色表示,可以提高地图的表现能力,使人容易发现数据的错误。编辑状态与制图状态分离,使某些在制图状态下发现的问题,必须返回编辑状态下修改,有时难以迅速找到准确的位置。使得单一工作流程出现多次反复,降低了工作效率,给制图者带来诸多不便。

4.2 MapCad 的制图特点

MapCad的基本制图模式如图4。

MapCad对几何信息的输入方法与一般GIS大同小异,但对属性数据,它输入的不是空间要素的属性码,而是对属性特征进行图形描述的参数或者说是图形属性码。

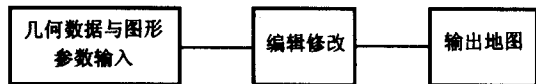


图4 MapCad的制图模式

Fig.4 Mapping scheme by using MapCad

MapCad的这种制图模式优点是:

制图过程简单,图形输入容易,建立图形快速

MapCad是采用图形属性(符号的形式、尺寸、结构、颜色等)、坐标和图元关系来描述空间对象,它是面向图式符号的或面向地图表示的,因而它独立于空间实体的具体含义。这种信息的存贮与它们的可视符号联系起来,使得制图数据结构简单,制图过程直观,即使一般制图人员经短期培训后也能迅速掌握系统的使用,因为对图形参数的理解比空间属性及关系的理解简单,直观得多。只要有地图图式、制图规范或者规定了合适的图形属性值,在输入地图位置坐标的同时,输入图形属性,通过编辑修改就能迅速输出地图。这种制图系统尤其适合制作由领域专业人员完成了作者原图的彩色地图的制作与输出。

编辑状态与制图状态的统一

MapCad以图形属性描述客观实体性质上的差异。编辑时屏幕上显示的图形已以相应的符号(线型、图案除外,目的是为了提提高显示速度)、颜色、尺寸和角度等反映,即编辑状态与制图状态融合,使制图者容易发现错误,实时修改错误,协调要素之间的关系。而不像一般GIS,要频繁地进行编辑状态和制图状态的切换。

该系统对地图制图最大的限制是:

(1) 不容易按属性查询修改和进行必要的空间分析。更新地图或编制新图需要对属性数据库进行操作,更新数据,提取信息。MapCad没有建立和管理属性库的功能,只能靠删除图面要素、更改图形属性值等对地图内容逐一处理。因此在原数字地图基础上更新地图或编制新图时,效率十分低下,不能充分发挥计算机系统快速更新数据的优势。

(2) 它不能与GIS中的数据进行有效的交流。没有属性库,使MapCad与GIS的结合产生困难。为了克服这一缺憾,新版的MapCad在输入时专门增加“扩展属性”功能,以此和GIS属性库联系。

5 地理属性与地图图形耦合的信息编码模型

通过对有代表性的软件ARC/INFO和MapCad的制图特点比较,可以看出两种不同编码思想对制图过程产生的影响。为了集中两类编码方法在制图上的优势,在此提出基于两种方法混合的编码模型——地理属性和图形属性耦合的信息编码模型(Coupled Geographical-Cartographic Attribute Encoding Model-CGCAEM),即同时用地理属性码和地图图形属性码描述空间实体的属性信息,以坐标(地理坐标和笛卡尔坐标等)表示地理实体的空间位置,以拓扑结构反映实体的空间关系。CGCAEM码位的逻辑构成如图5。

对于地理属性信息的编码,仍然遵循地学分类系统的拟定和编码原则。对于同一专题内容,其代码的位数相对比较确定,对于不同的专题内容,代码的位数可以不同。例如,对于划分到3级类型的土地利用分类系统,其地理属性码的位数可以是3位;而对于旅游资源,在“景系、景类、景型”自上而下的类型体系中再加上基类景型的“景域、景段、景元”3级规模体系,那么旅游资源的分类分级系统的地理属性编码码位至少为4位。

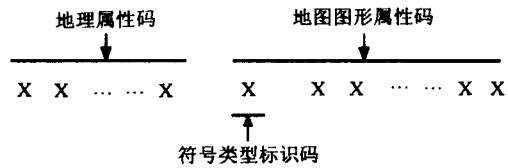


图5 CGCAEM 码位的逻辑构成

Fig.5 Logic Composition of CGCAEM fields

对于一个确定的 GIS 或 MS 而言,其地图图形属性的编码位数是固定不变的,基本上由地图符号的颜色、形状、尺寸、角度、纹理等基本图形要素组合而成。但是,哪些基本图形要素在一起是最佳组合,目前研究得较少,尚无统一结论。例如,下面是两种软件对线状要素图形属性位数与含义的确定:

软件 A: 线型(线的形状);颜色;线宽。

软件 B: 线型;辅助线型;颜色;辅助颜色;线宽;X 方向比例系数;Y 方向比例系数¹⁾。

显然对于同一个地图要素,如果图形属性码位不同,用户在产生和使用符号时操作和理解都不一样。

虽然地图符号千变万化,但是根据 A. H. Robinson 的研究,有一点应是明确的,即图形属性码的位数与含义应随着符号的类型(点状、线状和面状符号)的不同而变化。因为符号类型不同,构成其整体符号的基本要素不同。例如,内部填充了符号的面状图斑,可以看作是由多个点状符号有规律地排列而成,显然构成面状符号与独立的点状符号的基本图形要素是不一样的,因而这两类符号的图形属性也不一样。因此,在地图图形属性的编码构成中,应该有一位识别符号类型的代码,对符号进行分门别类,简化编码。

以 CGCAEM 方式记录的数据块中至少应包括

- 标识信息: 系统标识、用户标识、…;
- 地理属性信息;
- 图形属性信息;
- 控制信息: 输入控制、显示控制、…;
- ……。

在数据块中加入控制信息,主要考虑到(1)用户输入地图的不同目的和习惯,系统可提供以 GEEM、CEEM 或 CGCAEM 任一方式输入属性数据的功能,用户可按需要自由选择。如要快速建立地理数据库,可选择纯属性码输入方式;如要快速生产地图,

可选择纯图形属性码输入;或同时输入地理属性码和图形属性码。(2)加快显示速度。当所有符号均以“全真”模式显示时,速度会变慢。设置显示控制,用户可自由选择或取消全真显示模式,加快工作速度。

使用 CGCAEM 可以使得 GIS 在制图上具有更大的灵活性,主要有如下优点:

(1) 容易实现制图状态与编辑状态的统一。

(2) 容易更新地图和派生新图。空间实体的变化可表现为 3 种可能的形式^[11], (a) 属性变化,位置不变; (b) 属性不变,位置变化; (c) 属性和位置同时变化。这些变化有些涉及修改图形,有些仅涉及属性的修改,有些二者都要修改。由于图形符号与地理属性和空间坐标有机结合,很容易进行三者间的相互查询,这样就便于修改、编制或派生新图。

(3) 便于对有特殊显示要求的空间属性可视化。例如,在军事指挥、规划管理等方面的制图中,会遇到属性相同但在某一区域内要着重突出显示的情况。一般 GIS 只能将其处理成不同的属性来解决,但采用 CGCAEM,只需改变局部图形属性即可。

当然对地理属性和图形属性的耦合,会使得编码变得复杂,数据量增大,增加数据管理的难度。但是对于确定的 GIS 和 MS,由于图形属性编码位数固定,它们可以在地图要素属性对话框中显式表示,不会增加用户记忆的困难。随着科学技术的发展,计算机软硬件发展很快,因这种耦合而造成的数据量增大,不会给系统的存储、管理和使用带来太大问题。

为了使 CGCAEM 发挥更大的效用,必须强调图元库的建立,这也是加强 GIS 制图功能的途径之一。图元库可以包括点状图元库、线状图元(线型)库或者面状符号图元库等,主要反映基本图形要素中符号的“形状”。这里,图元可以是构成单一符号的基本图形要素,也可是构成整体地图符号的基本单元,甚至是复杂的象形符号。例如地图符号“●”,虽然由二个基本图形要素“○”和内部的颜色构成,但它既可以表示居民地,又可以表示景点,还可以表示地震震中,赋予不同的颜色,其含义更广,那么该符号就可以做为基元符号存放在图元库中。对于表示诸如景点的象形符号,难以分解,故可做为一个独立的图元。这样不论图元多么复杂,在编码上只用一位数码即可反映符号的形状。为了满足制作和管理图元的要求,应该强调建立、编辑、更新和管理图元库功能的设计与开发。利用这些功能,用户可以随心

1) 中国地质大学计算中心. MapCad3.0 微机编图系统使用手册.

所欲地去设计和制作符号。如果不具备这些功能,即使能建立起图元库,它也是一个死库,必将限制制图者的创意。

地理属性与图形属性耦合的编码方式,将使得MS与GIS进一步融合。GIS脱胎于计算机辅助地图制图,它与地图制图系统有着必然的联系,但也有本质上的区别。它们处理的对象都是空间实体,同样具有输出、存储、分析、制图和输出功能。在系统的结构、地图制作的方法与过程、硬件配置等方面,它们均可相同,有时确实难以区分,但是它们工作的重点不同,MS以短周期、高精度、低成本,专门生产地图为目的;GIS则着重于空间数据的管理与分析,地图仅仅是其信息表达的手段之一,GIS涉及的领域和用途更为广泛。因此在制图这一点上,许多GIS的功能反而不如MS专而精。进一步完善GIS中的制图功能,增强制图的灵活性;加强MS的空间分析与数据管理能力,进一步拓展计算机地图制图领域,是系统发展的要求。一次投资,多方多时受益,低投入,高产出,是用户的普遍需求。因此MS与GIS的融合应是今后发展的趋势¹⁾。地理属性与图形属性耦合的编码思想,属于为这种融合的可能性进行的基础性探索。

在研究工作中,作者就这一问题与同事同行们进行了讨论。“九五”国家重中之重科技攻关项目“地理信息系统基础软件的开发与商品化”支持的CITYSTAR软件,在1996年的总体与功能设计中采

纳了作者文中的思想。

参 考 文 献

- 1 周成虎. 地理信息系统的透视——理论与方法. 地理学报, 1995, 50: (增刊).
- 2 陈述彭. 地理信息与地理信息系统. 地理学报, 46(3).
- 3 P. A. Burrough. Principles of GIS for Land Resources Assessment.
- 4 黄杏元, 汤勤. 地理信息系统概论. 高等教育出版社, 1989.
- 5 马藹乃. 地理遥感信息模型. 地理学报, 1996, 51(3).
- 6 田德森. 现代地图学理论. 测绘出版社, 1991.
- 7 高俊. 地图的空间认知与认知地图学——地图学在文化与科技领域的新探索. 第四届全国地图学学术讨论会论文集. 中国地图出版社, 1992.
- 8 焦健等. 地图出版系统在专题制图中的应用分析. 地图, 1995, (3).
- 9 A. H. Robinson, et al. Elements of Cartography (5th ed.). John Wiley & Sons, Inc, 1985.
- 10 ESRI. Understanding GIS-The ARC/INFO Method.
- 11 陈军. GIS空间数据模型的基本问题和学术前沿. 地理学报, 1995, 50: (增刊).

作 者 简 介

焦健,女,1960年12月生,副教授。1983年毕业于武汉测绘科技大学地图制图专业,1988年在北京大学遥感研究所获硕士学位,同年留校任教至今。1996年至1997年在法国普鲁旺斯大学地理信息处理中心做访问学者。现从事的领域为计算机地图制图、GIS和遥感应用。发表论文和地图作品近20篇(幅)。

Coupled Geographical-Cartographic Attribute Encoding Model

Jiao Jian Zeng Qiming

(Institute of Remote Sensing and GIS, Peking University, Beijing, 100871)

Abstract Spatial information encoding which presents the interface between the real and information entities, is one of the most important steps in designing GIS. Most GIS and mapping systems almost adopt the ways of encoding for geographical attributes, which are more effective for spatial analyses but weaker in mapping. Referring to the other coding methods such as AutoCad and MapCad, the authors raised the concepts of geographical entity encoding model (GEEM) and cartographic entity encoding model (CEEM), and compared their advantages and disadvantages in mapping. A new model, coupled geographical-cartographic attributes encoding model (CGCEM), has been proposed based on the two models mentioned above.

Key words Information encoding model, GIS, Mapping

1) 焦健等. 融于GIS的专题制图系统功能设计中的几个问题. 中国GIS协会首届年会论文集, 北京, 1995年.