

计算机辅助地质填图属性数据采集子系统的 动态数据模型

张夏林 汪新庆 吴冲龙

(中国地质大学资源学院, 武汉 430074)

摘要: 对于复杂的地质属性数据, 使用常规的数据库数据模型, 需要多界面录入数据, 要耗费大量时间在界面的切换上, 培训使用系统困难; 另外, 系统完成封装后, 模型和界面就被固定, 无法做必要的实时扩充或更改。使用数据字典技术、数据逆向规范化技术和数据挖掘技术支持下的动态数据模型, 并把数据录入模型与数据存储模型分离可以很好地解决以上问题, 提高了系统的实用性、适应性和生命力。

关键词: 数据模型; 动态数据模型; 数据库; 地质填图; 计算机辅助地质填图系统。

中图分类号: P628⁺.4 **文献标识码:** A

文章编号: 1000-2383(2001)02-0201-04

作者简介: 张夏林, 男, 1975 年生, 中国地质大学(武汉)在读博士生, 主要从事信息系统的研究开发。

计算机辅助地质填图系统的设计, 采用地矿点源信息系统的原理与方法^[1,2], 以掌上机、袖珍便携机、PC 微型机和 GPS、DBS、GIS 等为硬件和软件支撑。其中, 野外数据采集子系统可分为空间数据采集和属性数据采集两大模块。属性数据采集模块主要使用数据库技术, 一步到位地将野外属性数据以规范的格式录入计算机, 支持后期各种应用。为此, 建立一个完善的数据模型对保证采集信息的完备性、规范性和一致性是至关重要的。

1 常规数据模型的局限性

地质属性数据具有数据源丰富、数据量庞大、数据类型众多、数据结构复杂的特征^[1]。所谓数据结构复杂, 就是指数据具有多层次性(多重父子关系)。例如在区域地质调查工作中, 所要收集的地质信息谱系^①如图 1 所示。图中每一个实体的属性表集可能包含多个层次的属性表。例如, “地层”实体的属性表

集就包含了 3 个层次的属性表: 地层基本信息、地层化石和地层综合信息。每个属性表又包含多项属性。对于这样复杂的信息, 使用常规的数据模型, 就意味着一线的区调人员必须熟悉数十个乃至上百个数据表结构及其输入界面, 这样, 不仅系统操作培训需要花费很长时间, 而且在野外使用时也要耗费大量的时间在界面的切换上。另一个更为突出的问题是, 如果使用常规的数据模型, 一旦系统开发完成封装后, 模型和界面就被固定了, 无法更改。而在现实的地质填图工作中, 野外人员常常希望根据工作区的性质自主地调整数据模型。由于信息的复杂性, 对于填图考虑再全面的数据库系统, 数据模型都会有不能满足需要的时候。能否构建一个具有普适性和可扩充性的野外数据采集系统, 这是众多系统开发单位的追求和野外工作人员的期盼。动态数据模型及其相应技术的使用, 可以实现这一理想。

2 动态数据模型的概念及实现方法

动态数据模型是相对于常规数据库模型的一种在系统封装后, 允许使用者在使用系统过程中, 根据需要实时地增减数据表或数据表中字段的数据模型^[3,4]。动态数据模型也是一系列规范的数据库表,

收稿日期: 2000-11-20

基金项目: 国土资源部 95002 项目; 中国地质大学“数字地大”项目。

①吴冲龙, 汪新庆, 刘刚, 等. 计算机辅助 1:5 万区域地质填图系统研发报告. 2000.

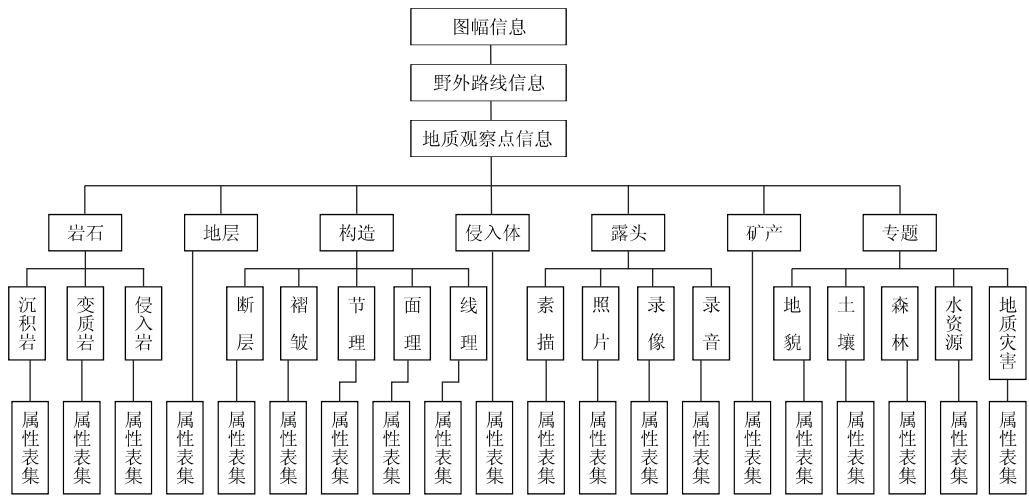


图 1 野外地质属性数据谱系图

Fig. 1 Pedigree chart of field geology attribute data

它首先由系统设计者根据区调填图的实际工作数据模型,抽象出基本能满足各种性质工作区和各种工作目的的一个基本数据库模型,作为支持系统的“原型”模型。这个“原型”模型跟普通的数据库模型没有太大的差别,仍然要遵守数据库模型的各种规范原则^[5],但它不同于普通数据库模型,它是一个开放的、未定型的模型,在使用中可以实时修改、增删而不会影响系统正常工作,能够对使用中所更改的模型及录入的数据做常规处理。

2.1 数据字典技术

在关系数据模式设计与物理模式设计中,数据字典是实现数据库的安全性、完整性、一致性、可恢复性、有效性、可修改性以及可扩充性的重要手段^[1,3]。数据字典所维护的数据称为元数据,即用两个系统数据表存储和管理整个数据模型。利用数据字典,可以使系统实时跟踪和记录使用者对数据模型的更改,并对更改做出响应。数据字典的样式如表 1、表 2 所示,用户每次对模型的修改,都被保存到以上字典中,在录入界面上显示录入项目,系统先到字典中读取最新的值,保证用户所做的修改被及时响应。

2.2 数据录入模型与数据存储模型分离

通过数据字典,模型可以动态地被修改,但是,修改后的模型仍然是规范的常规数据库模型,如果录入界面直接操纵这些数据表,地质人员为了录入关系复杂的地质数据,还是要陷入大量的录入界面选择和切换的繁琐操作中。此外,对于封装好的系统,目前还难以做到动态增减数据录入界面上的元

表 1 数据表字典

Table 1 Data table dictionary

字段说明	字段名	类型	长度
通用编号	main_code	字符型	2
实体编号	sub_code	字符型	2
实体名称	sub_name	字符型	24
数据表名	main_file	字符型	20
表说明	table_capt	字符型	48
表路径	table_path	字符型	32
索引名称	tag_name	字符型	10
主关键字	first_key	字符型	10
关键字名	key_name	字符型	16
数据库名	db_name	字符型	8
选择控制	select	逻辑型	1

表 2 表字段字典

Table 2 Field dictionary

字段说明	字段名	类型	长度
字段名称	field_name	字符型	10
字段说明	field_capt	字符型	32
通用编号	main_code	字符型	2
实体编号	sub_code	字符型	2
字段序号	field_num	字符型	2
保留字段	field_show	字符型	4
选择控制	xz	逻辑型	1
是否必需	requisite	逻辑型	1
唯一性	onlyone	逻辑型	1
标志	lsign	逻辑型	1
优先性	grand	逻辑型	1

素。为了支持动态数据模型,并给用户一个简洁的数据录入界面,我们采用了数据录入模型与数据存储模型分离的方法。其中,录入模型使用抽象数据模型。

为了保证数据所在相对父子关系的正确,可先将可能的信息按其所处父子关系分为 4 个层次:(1)图幅级;(2)路线级;(3)地质点级;(4)细节信息。前三级作为父表相对稳定,所以采用常规数据模型,第四级因其模型多变、数量大,使用抽象数据模型(表 3)。该模型中把所有可能的属性数据抽象为三项:第一项是描述项目,或称为数据项,即地质人员要从哪些方面描述一个地质体或一种地质现象。每一个实体有一套基本的描述项目,按不同类型保存在数据字典中,用户可根据需要任意选用,效果等同于动态的数据模型。第二项是描述值,或称为数据值,即地质人员选定一个描述项目后,在这一项中选择输入该描述项的值。一般采用国家标准,由系统提示输入^[6]。第三项是自由描述,即地质人员对描述项目的补充描述或地质认识,可以根据需要定义其长度,这样就让地质人员充分描述感兴趣的内容。例如描述项目为“GZEEC(断层力伴生构造类型)”,描述值为“GZEZZ(牵引褶皱)”,自由描述为“该牵引褶皱的突出方向指示断层本盘的运动方向”。通过使用此抽象数据模型,地质人员所有可能要描述的内容都可以用一个简单的界面快速录入到数据库中。

2.3 逆向数据规范化技术

数据规范化是指依照关系数据库理论,按范式分解的方法把实体的属性集分解为一组结构简单、逻辑严密的二维平面表的过程^[5]。然而,为了使用上述的抽象数据模型,需要一种逆向数据规范化技术,即把动态模型中那些已经规范化的数据表,转化为表面看似不规范的抽象数据模型的数据表。实现逆向数据规范化的关键是采用数据字典技术,地矿点源数据库计算机辅助设计(GDCASE)技术^[1,3]提供了这方面的借鉴。

2.4 数据挖掘技术

存储在抽象数据模型(表 3)中的数据不符合数据库原则的数据,不能被直接使用,必须将其再次按规范的数据格式提取出来,转存到存储模型中。这个过程属于数据挖掘,由专门的功能模块执行。处理程序根据抽象数据模型中的关键字,将暂存在抽象数据模型表中的数据,规范地转存到数据存储模型的数据表中,并保证信息无损。此过程对地质人员是透明的,即地质人员并没有感觉到抽象模型的存在。动态数据模型作为数据存储模型,数据采集到一定程度后,处理程序将暂存在抽象数据模型表中这些数据转存到数据存储模型的数据表中,得到的记录

表 3 抽象数据模型

Table 3 Abstract data model

字段说明	代码(国标)	类型	长度
地质点号	qdebga	字符型	6
通用编号	main_code	字符型	3
描述项编号	item_code	字符型	3
描述项目	items	字符型	10
描述值	item_value	字符型	10
自由描述	ctext	字符型	120

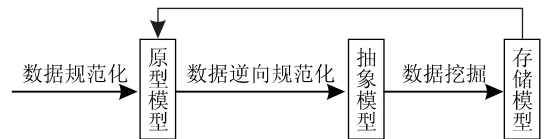


图 2 数据模型转换示意

Fig. 2 Sketch of data model conversion

集为常规的规范化数据库数据记录集。整个过程如图 2 所示。

3 动态数据模型利弊分析

(1)动态数据模型的优点。①动态数据模型允许野外工作人员在保证信息完备的前提下,针对具体情况订制实用的模型,使系统具有了普适性,增强了计算机辅助填图系统的适应性、实用性和生命力。②增强了计算机辅助填图系统的易使用性,可以提高工作效率。采用动态数据模型和抽象数据模型,系统可以使用唯一而简洁的数据录入界面(图略)采集全部的数据。这个录入界面的使用非常方便,野外人员的操作培训时间可由数周缩短为几小时。

(2)动态数据模型的缺点。首先是要求系统在运行中能修改数据库的数据模型,这就需要解决一系列相关的技术问题,才能保证模型的有效性和正确性;其次,使用了动态数据模型,系统灵活了,但系统也处在一种相对不稳定状态;最后,与之配套的抽象数据模型的使用,由于违反了数据库的数据规范化准则,会出现一定的数据冗余,增加存储空间的消耗。

4 结语

动态数据模型是解决使用野外属性数据计算机辅助采集模块的许多困难问题的重要技术措施,是提高系统的生命力和适应性、促进系统走向实用化

的重要途径。尽管使用动态数据模型会产生一些新问题,但是这些问题是可以避免、消除或妥善处理的。

参考文献:

- [1] 吴冲龙,汪新庆,刘刚,等.地质矿产点源信息系统设计原理及应用[M].武汉:中国地质大学出版社,1996.
- [2] 吴冲龙.地质矿产点源信息系统开发与应用[J].地球科学——中国地质大学学报,1998,23(2):193~198.
- [3] 汪新庆,刘刚,袁艳斌,等.岩土工程勘察点源程序信息系统的开发[A].见:岩土工程青年专家学术论坛文集编

委会.岩土工程青年专家学术论坛文集[C].北京:中国建筑工业出版社,1998.73~78.

- [4] 汪新庆,刘刚,韩志军,等.地质矿产点源数据库系统的模型库及其分类体系[J].地球科学——中国地质大学学报,1998,23(2):199~204.
- [5] 冯义才.数据库系统基础[M].武汉:华中理工大学出版社,1993.
- [6] 汪新庆,刘刚,袁艳斌,等.地质矿产术语分类代码在地矿点源信息系统中的应用[J].地球科学——中国地质大学学报,1999,24(5):529~532.

DYNAMIC DATA MODEL APPLIED IN REGIONAL GEOLOGICAL MAPPING COMPUTER-AIDED MAPPING SYSTEM

Zhang Xialin Wang Xinqing Wu Chonglong

(Faculty of Earth Resources, China University of Geoscience, Wuhan 430074, China)

Abstract: When we construct a database to store the complicated geological attribute data, we need many interfaces to input the data if we use the routine database data model. So it will be very difficult to input the data into right places and learn to use this kind of system. In addition, it is impossible to do some necessary change and expansion to the data structure and interface as soon as the system development is completed and the data model and the system interface are fixed. These problems can be solved by using the dynamic data model, which is established based on the data dictionary technique and data converse standardization technique, data digging technique and by separating the inputting data model and data storage model. As a result, the database built by using the dynamic data model can be more practicable, adjustable and vital to the system.

Key words: data model; dynamic data model; database; geological mapping; computer-aided mapping system.