(19) 330-335

农业生态研究文档数据库的概念模型设计

施建平 王德建
(中国科学院南京土壤研究所 南京 210008)

S181
S126

摘 要 农业生态研究数据含有随时间、空间变化的信息,其数据类型复杂,且数据类型更新频繁。原有的商用数据库系统不能满足数据管理的需要。从用户查询角度,建立以实体-关系(E-R)模型为基础的农业生态研究文档数据库概念模型,能够管理地图数据、统计数据和野外观测试验数据,满足多种数据类型的管理和查询的需要。本文介绍了实体-关系概念模型的设计,以及根据概念模型建立的农业生态研究文档数据库。

关键词 农业生态;数据库;概念模型

农业生态研究数据不但含有随时间变化的信息,而且含有地域性的空间信息。用于农业生态研究的数据应包括:来自于地图或遥感数据的地形、地貌、土壤、土地利用、植被的向量或栅格数据;来自于统计网络的气候统计和社会经济统计数据;还包括野外观测试验数据。以便根据野外长期观测试验点的经验和数据,定性或定量地进行 GIS 区域性分析,实现研究成果的空间拓展。

随着农业生态方面数据需求的日益增长,数据的组织和管理已成为重要的议题。关系型数据库在 80 年代被确立后,随着不断的成熟和发展,至今已成为数据库管理系统的主流。关系型数据库是应用二维表来表示和处理信息的实体集合和属性关系的数据库。它并不是按物理的存储方式来组织连接数据,而是通过建立表与表之间的关系来连接数据库中的数据。实体(Entity)是我们关心的、存在于客观世界中的、并要记录和加工的信息对象。关系数据库中用表描述某一实体。表由记录和字段组成。表之间相互独立又相互联系。表的独立性是通过"实体"实现的。表的相互联系是通过关系(Relation)实现的。实体-关系模型定义了一个关系型数据库的结构。

国外许多农业生态环境数据库大多采用实体-关系模型作为其数据管理模型[1~3]。国内 90 年代初期,中国生态系统研究网络制定了生态站历史数据整编元数据(metedata)标准,用于规范现存的生态研究数据。由于当时的特定条件所限,上述数据是基于文档文件管理,缺乏便于查询的关系数据管理功能。

为了更好地组织管理数据,我们参照美国联邦地理数据委员会(FGDC)空间数据的元数据标准①,应用实体-关系模型设计了农业生态研究数据管理系统的概念模型。首先将分散的数据按照应用范围或类型(地图、文子、表格、图像)规范建立文档,然后依据数据文档的专题分类、数据获取时间及数据获取地点查询和管理数据,建立文档数据库。在设计中考虑到以下几点:

① Federal Geographic Data Committee, 1994. Content Standards for Digital Geospatial Metadata. (ftp://waisquarse.er.usga.gov/wais/docs/FGDCmeta52394.ps)

1. 为了保证数据的科学价值，对长期实验数据或特殊的非长期实验数据建立文档说明，特别说明其研究目的，实验设计和实验方法。即使数据的提供者离开该课题或课题结束，数据始终是可用的。

2. 良好的用户界面。在用户查询具体数据之前，需要了解该数据集在何时、何地、为何目的而采集的。应具备按时间、地点、专题进行查询的功能。用户无须了解数据库内部结构，可直接根据数据库提供的用户界面查询数据。

3. 适于管理各种复杂的生态数据类型(地图、影像、野外观测实验、关系数据表格、图形、文本)。

4. 对于较大数据集(如GIS数据)，数据库不存储检索原始数据，仅提供数据集的获取路径，存取方法，空间数据的投影和比例尺等元数据。

5. 具有较为灵活的用户界面，能与其他软件方便地交换数据，便于模型统计和计算。

根据以上考虑和太湖地区农业生态研究实际需要，我们设计了农业生态研究数据管理系统的概念模型，并应用Oracle开发了常熟农业生态研究文档数据库，其主要目的是：提供一个面向用户、便于查询使用的数据库管理系统。

1 概念模型设计

在设计概念模型时首先要了解数据所反映自然现象的本质，农业生态研究数据本身含有地域性的空间信息(如不同地点的肥料试验)，不同层次的分类信息(如土壤分类关系)。其次要考虑为了查询和管理数据方便而形成的各种特殊关系，如不同数据类型的分类信息，查询项目与数据集的关系，数据集与获取方式的关系等。因此，从空间分布关系、分类关系和查询关系角度设计概念模型。

1.1 空间分布关系

包括野外观测试验、生态学监测、农田小气候等环境因子和社会经济统计的数据集，均含有空间分布的特征，仅分别用单一的关系数据表格来描述不能反映自然现象的本质。例如，一个肥料试验可能涉及到不同地点的土壤类型，一个数据集可能涉及到多个地点。反之，一个地点(如常熟农业生态站所在地常熟市辛庄镇)，也可能有多项观测试验进行。在GIS系统中用点数据描述上述肥料试验数据的二维空间分布关系。农业生态学数据还可能有三维空间分布的关系。例如，一个土壤类型有多个地点的空间分布；有多个采样点和不同的剖面分布；每个剖面有多个层次及其对应的理化特性。农田小气候观测数据也具有三维空间分布特征，例如大气温度的数据可能来自不同的观测场地和不同的观测高度。

1.2 分类关系

农业生态学研究的数据集可以是地图数据、遥感数据、野外观测试验或文本数据。为便于管理，将数据类型分类为图像数据(地图或遥感数据)、关系表格(野外观测数据)及文本数据，采用不同的存储及查询方式。

图像数据：由于数据库软件本身的限制，不能用现有的机器管理占据巨大存储空间的地图或遥感数据，但设计查询其元数据(包括比例尺，投影，说明，压缩图像)，用户通过查询元数据和浏览压缩图像，了解数据的特征。由数据的存取说明，或html链接指向获得数据集。

关系表格：观测试验数据大多由关系数据表格组成。观测试验数据元数据应包括观测

名称、数据类型、采样频率、研究目的、试验设计和方法、试验仪器等说明。用户在获取数据前首先了解形成数据的研究目的和实验方法。

文本或管理数据：文本数据来自重要的文档文件，例如成果说明、中英文简介、有关的野外站区图、课题汇总、人员名单等。其元数据应包括：标题、摘要、是否发表或获奖、发表(获奖)日期、期刊等。文本数据在桌面数据库和 Oracle 数据库中均可用不限长度的字段表示。

在概念模型设计中，首先提取数据集公共的信息作为数据通用信息。设计一个表格"数据集基本信息"，其字段包括：数据集名称、来源、数据集类型、数据起始时间、数据结束时间、数据的提供者、数据提供组织、数字化状态等。然后，分别提取地图数据、野外观测数据和文档数据的元数据，组成不同的分类数据表格："专题图"、"观测数据"、"文本数据"。由于一个数据集可能存在多个数据子集，专题图与分类数据表格通过数据集标识符"数据集-ID"相连接，其关系为一对多的逻辑关系。

1.3 查询关系

考虑用户一般性的查询需要，设计按数据采集时间、采集地点和专题查询的功能进行。农业生态数据涉及空间数据范畴，每个数据集可能含有多个数据采集地点，某个地点可能涉及到多个数据集。数据集与数据采集地点为多对多的关系。例如，常熟社会经济统计数据提供了以乡为单位农村社会经济统计数据，包括常熟辛庄，大义等 32 个乡镇。而常熟辛庄，作为中国科学院常熟农业生态实验站所在地，同时还收集其它试验研究数据。同理，一个数据集可能属于多个专题范围，一个专题词可能包含多个数据集。例如，野外观测"水分状况对水稻土中物质迁移及作物生长的影响"数据集，可以属于"土壤"和"农业生态学"两个专题范围；"土壤"作为一个专题关键词，可以有多个数据集引用。数据集与专题关键词同样为多对多的关系。

表 1 专题分类查询表格间关系

数据集基本信息

数据集-ID	数据集名称
1	土壤空间数据库
2	社会经济统计数据库
3	常规气象观测
7	土壤肥力监测
8	水分状况对水稻土中物质迁移及作物生长的影响

专题索引

数据集-ID	专题-ID
1	1
7	1
8	1
2	2

专题

数据集-ID	专题名称
1	土壤
2	社会经济
3	气象

其次，分析数据集与存储方式的关系。为安全起见，一个数据集可能有多个的存储地点和备份文档。数据集可能存放在不同的存储介质上。已数字化的数据集可能存放在网络上多个计算机、服务器、或备份磁带、磁盘上。未数字化的数据集以地图、出版物、研究报告及其拷贝存放在多个地点。而一个存储介质(如数据流磁带)又可能存放多个数据集。数据集与存储方式为多对多的关系。

如前如述，由于用户查询需要，数据集与数据采集地点、专题、存储方式为多对多的逻辑关系。而实现多对多逻辑查询对于一个关系数据库来说实现查询非常困难。解决这一矛盾的方法是：设计一个介于两个多对多关系中间的表格，将多对多关系转换为多对一和一对多的关系。例如，为了解决数据与专题关键词多对多的关系，首先将数据集一般性信息建立表格"数据集基本信息"，"专题关键

词信息建立表格“专题”，再建立一个中间表格“专题索引”。“专题索引”联接“数据集基本信息”表格与“专题”表格，它由两个字段“数据集-ID”和“专题-ID”组成，描述两者之间一对多或多对多的关系。

例如，数据集“水分状况对水稻土中物质迁移及作物生长的影响”，可以属于分类“土壤”和分类“农田生态学”两个专题范围；“土壤”作为一个专题关键词，可用于数据集 1“土壤空间数据库”、数据集 7“土壤肥力监测”和数据集 8“水分状况对水稻土中物质迁移及作物生长的影响”（表 1）。

1.4　复杂数据集概念设计实例

1. 土壤数据。主要来自土壤普查数据的土壤数据集，由于它的空间分布特性和分类的层次关系，单用一张二维表格不能表示其自然特征，需要用多个数据表格表示。一个土壤类型有多个地点的空间分布，在土壤图上用多个图斑表示。每个土壤类型有其典型的剖面分布，每个剖面有多个层次及其对应的理化特性。我们用 4 个二维表格来描述其复杂的空间分布关系。首先建立表格“土壤”描述土壤类型一般性的信息，包括土壤类型标识符“土壤-ID”、“名称”、所属“土类”、“亚类”。其次，建立表格“土壤多边形”连接，表示图斑拓扑关系的多边形及其对应土壤类型。“土壤”与“土壤多边形”通过土壤类型标识符“土壤-ID”相连接。一个土壤类型属性可以有多个多边形图形表征其空间分布，“土壤”与“土壤多边形”之间为一对多的关系。其次，建立描述景观特征的表格“景观”。它包括景观标识符“景观-ID”、“高程”、“母质”、“土地利用信息”。一个土壤类型可能有多个景观分布，表格“土壤”与表格“景观”间为一对一或一对多的关系，通过土壤类型标识符“土壤-ID”相连接。最后，建立剖面理化性质表格“剖面”。“剖面”描述特定剖面不同层次的名称、层次厚度、颜色、及对应的物理化学性质。表格“景观”通过标识符“景观-ID”连接表格“剖面”，并反映景观与剖面理化性质一对多的关系。此外，在表格“土壤”中还增添土壤类型分布的浏览图像字段，以便查询时直观表示该土壤类型的空间分布特征。

2. 气象观测数据。农田小气候自动观测数据按每小时一次的采样频率采集储存数据。若研究积温和降水对作物的影响时，需要了解大气温度、土壤温度和降水量的月分布状况。因此，农田小气候观测数据需要按日、月统计，汇总为年报表、月报表。若每年建立一个年报表和一个月报表及一个原始记录文件，随着数据的累积将会给数据管理造成困难。因此，使用关系模型建立农田小气候数据子库。

首先，建立农田小气候观测年报表，它应包括数据集标识符“数据集-ID”，年，月，及按月统计的气温、降水、湿度，风速等气象要素。其次建立农田小气候观测月报表，它应包括与年报表对应的年，月，及日和按日统计的气象要素。显然，一年的数据中包含多个月的气象要素记录，而一个月的数据包含多个日的气象要素记录。年报表与月报表的关系为一个对多个的关系。

由以上对各种类型数据关系的分析，建立了关系－实体(E－R)数据模型（图 1）。

2　模型的应用

依据上述概念模型，应用 Oracle/Access 数据库软件，建立了常熟农业生态研究文档数据库系统。收集和建立了 11 个数据集及对应的数据子库。它们是：常熟市土壤空间数据库

(GIS数据库);常熟市社会经济空间统计数据库(GIS数据库);常熟市常规气象统计数据库(统计数据);常熟站农田小气候观测数据库(观测试验数据);土壤肥力监测数据(观测试验数据);常熟市环境污染数据库(统计数据);稻麦两熟农田生态系统碳素循环(野外试验数据);常熟农业生态站基本管理信息(文本、表格、图形);水份状况对水稻土中物质迁移及作物生长的影响(野外试验数据);中国科学院常熟农业生态站及其毗邻地区土壤图(地图);农业环境空间分布图形库(图形)。

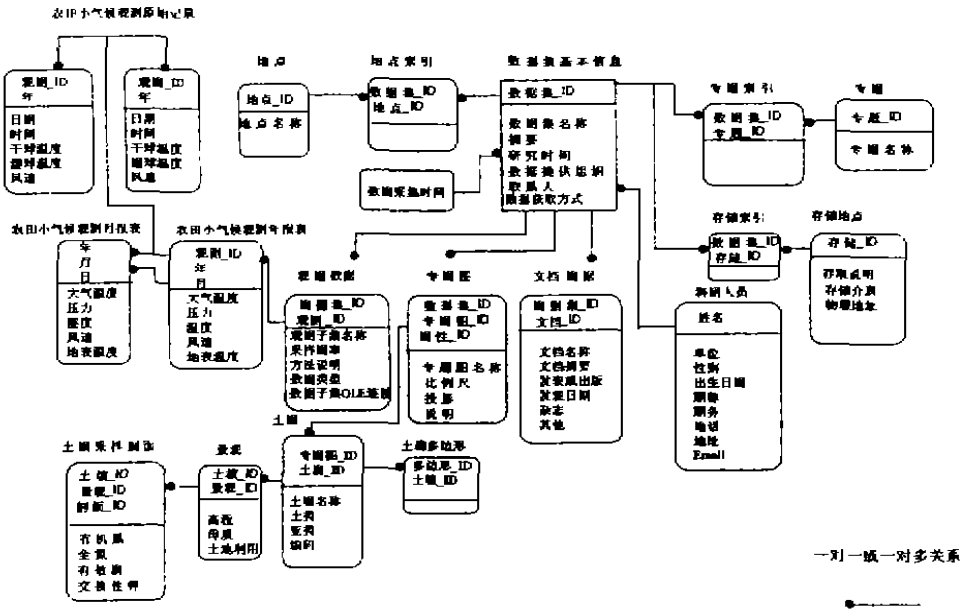

图 1 农业生态研究文档数据库概念模型

用户在寻找数据时,首先思考的是他所要的数据是在什么时间、地点被收集,属于什么专题范围。按照用户的需要,设计 3 个基本查询功能:时间查询、地点查询、专题查询。即:选择一个时间段,查看什么专题数据存在于数据库内,在什么地点被收集;选择一个地点,查看什么数据集在数据库中,在什么时间段被收集;选择一个专题,查看属于该专题数据集的采集时间和地点。以查询土壤数据库为例,说明查询过程。首先,从专题查询着手,查询关键词"土壤",检索到六个数据集与"土壤"有关。选择数据集 1"常熟市土壤空间数据库"。从数据集一般性信息中可得知,该数据来源于全国第二次土壤普查资料:常熟土壤志,数据采集时间为 1985 年。其次,选择摘要显示。摘要描述了土壤图的比例尺、投影、图编码及浏览图形。若用户对此数据感兴趣,则打开数据集子库,进行查询或浏览。常熟市土壤数据库是以常熟市土壤普查资料为基础的,选择土种为基本单元,以该土种的典型剖面数据为代表。一般地,用户希望选择一个土种,了解其空间分布、分类属性、景观、及理化特性。在查询界面上设计了土种选择窗口,当一特定的土种被选中后,数据库将在屏幕的主窗口显示该土种的分类属性、空间分布图像、地形、土地利用、剖面描述等信息,在子窗口显示该剖面的不同层次土壤理化特性(图 2)。

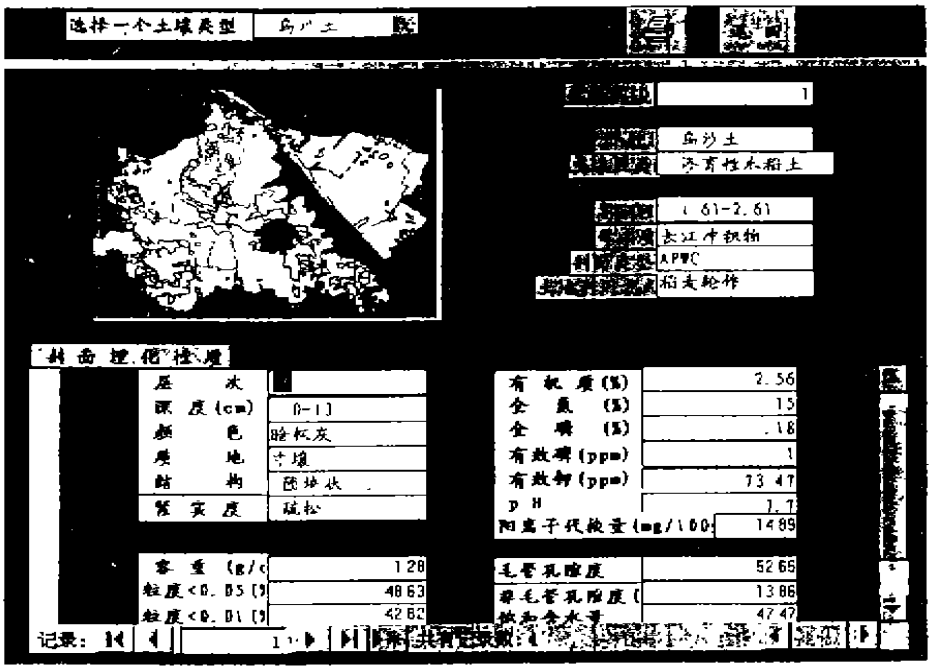

图 2　土种查询界面

3　结语

农业生态研究数据含有随时间、空间变化的信息，其数据类型复杂、更新频繁。原有的商用数据库系统不能满足数据管理的需要。从用户查询角度，建立以关系数据模型为基础的农业生态研究文档数据库概念模型，能够管理地图或遥感数据、统计数据和野外观测试验数据，满足多种数据类型的管理和查询的需要。以模型为基础、按照用户思维方式而建立的用户查询界面，较好地解决了面向用户的使用问题。采用基于 Oracle/Access 的客户/服务器方式，为今后网络分布式数据库的应用打下了基础。

对农业生态研究文档数据库的应用表明，关系模型作为农业生态研究数据管理的概念模型是可行的，能够满足大量的数据管理和查询的要求。应用关系模型建立的数据库管理系统，具有长远的生命力和广泛的应用前景。

参 考 文 献

1　P. A. Burrough. Principles of Geographical Information Systems for Land Resoureces Assessment. Clarendon Pess, 1986

2　L. R. olderman. In transaction of 14th International Congress of Soil Science. V. 1990, 136～140

3　陈俊，官鹏. 实用地理信息系统. 科学出版社，1998, 63～73