黄土高原资源环境数据库中科技文献摘录数据的管理

郭明航[1,2], 李够霞[1,2], 吴开超[3], 刘 峰[3]
(1. 西北农林科技大学水土保持研究所;
2. 中国科学院水利部水土保持研究所, 陕西 杨陵 712100; 3. 中国科学院计算机网络信息中心, 北京 100080)

摘 要: 通过对科技文献中的科学数据特性的分析, 采用 B/S 数据库体系结构以及合理的安全和用户管理策略, 应用 Oracle 数据库系统, 建立了面向用户的数据查询和用户自主管理数据的网络数据库管理平台。数据的选编突出黄土高原生态研究的区域特色和重大科技需求, 偏重生态因子、生态过程的试验、监测数据收编, 同时注重数据的基础性、客观性、历史性特点。对于时空跨度较大的生态研究来说, 本数据库无疑是一个具有历史价值的、专业特色突出的科学研究基础数据库。

关键词: 科技文献; 科学数据; 数据库; 黄土高原
中图分类号: TP392; X171.1 文献标识码: A 文章编号: 1005-3409(2006)02-0186-03

Management of Excerpted Data from Technological Literature in Loess Plateau Resource Environment Database

GUO Ming-hang[1,2], LI Gou-Xia[1,2], WU Kai-chao[3], LIU Feng[3]
(1. Northwest Sci-tech University of Agriculture and Forestry;
2. Institute of Soil and Water Conservation, Chinese Academy of Sciences and Ministry of Water Resources; Yangling, Shaanxi 712100, China;
3. Center of Computer Network Information, Chinese Academy of Sciences, Beijing 100080, China)

Abstract: Network database system with data inquire and consumer self-determination function were built by analyzing data characteristic from technological literature, using B/S database system configuration and its rational safety and consumer management strategy, and applied Oracle database system. Database excerpted from technological literature considered local characters and national technological requirement for ecological research in Loess Plateau, particularly stressed on data excerption about ecological factor, ecological process and ecological monitor. The database system provides with special and useful data for the ecological research with big temporal and spatial span.

Key words: technological literature; science data; database; Loess Plateau

随着人类进入信息时代, 科学数据已经成为一种重要的资源和社会财富, 它贯穿于科技活动的全过程, 渗透到人类生活的方方面面, 影响着整个社会的发展。所谓科学数据它是人类为了认识世界和改造世界而用于记忆世界的一种符号, 这种符号可以是数字、文字、符号、图形、声音、图像等等。而信息则被认为是物质的一种普遍的属性, 它反映不同物质所具有的不同本质、特征以及运动状态和规律。数据和信息的区别就如同原材料和产品一样, 人们通过加工原材料成为产品, 并被人们所使用[1]。简言之, 数据是信息的载体, 信息是数据内在属性的表示。科技文献是科学数据和信息的复合体, 科技文献在传播的过程中其数据和信息起着不同的作用。第一, 科技文献的主要目的是传播信息, 而数据则是隶属于信息, 对信息起支撑作用。第二, 相同的数据用不同的方法加工处理, 或者被不同的人加工处理可能产生不同的信息。这是因为数据本身所固有的属性所决定的, 数据的这些属性包括分离性、共享性、客观性、长效性、非排他性、增值性、传递性、资源性、公益性等[2]。第三, 科技文献所发表的数据一般都是是经过作者从大量的科研活动中选取的, 因而具有较高的可信度和科学性, 值得长期收藏。基于对科技文献中信息和数据的这种基本认识, 建立科技文献数据摘录数据库, 使之成为科学研究、经济发展、社会进步的有用资源, 并在全社会共享就成了本数据库建设的初衷。目标是解决从科技文献中摘录的科学数据的管理问题, 将各种形式的出版物公开发表的科学数据资源组织起来, 为科学数据的使用和科学数据库建设提供一个工具, 将数据资源与计算机、数据库和网络等先进技术相结合, 促进数据向知识的转化以及信息化的科学研究环境的建立, 并为科研与社会提供科技数据资源共享与服务。

1 数据库的设计与开发

1.1 数据库的管理对象

经过对科技文献摘录数据的综合分析, 可归纳出数据库的基本管理对象为以下三类:

(1)数据集。数据集是本数据库管理的基本单元, 一个数据集包括数据集基本信息、数据集实体信息。

[1] 收稿日期: 2005-12-26
基金项目: 中国科学院“十五”信息化项目, 科学数据库建设及应用, 项目编号: INF105-SDB-1-31
作者简介: 郭明航(1962-), 男, 高级工程师, 从事科学数据管理研究。

(2) 数据实体。数据实体是一个数据集的核心内容, 它是一系列拥有共同主题以及方法等特征的数据系列的集合, 也就是常说的数据表。

(3) 数据出处。数据出处就是说明数据来源的信息, 这类信息是为更进一步查阅数据实体而设立的参考信息。根据通用的做法, 数据的出处分为专著、期刊、论文集、报告、学位论文、通用类型、电子文献类型等 7 种。

1.2　数据库的功能设计

科技文献数据摘录数据库是基于 B/S 架构的数据库系统。数据库软件选用 Oracle9i, 应用软件采用 Java 语言开发。数据库的基本功能包括:

(1) 数据录入。提供友好的数据著录功能及其界面, 实现对数据的添加、删除、修改等数据库基本操作。

(2) 数据查询检索。提供友好的数据查询功能及其界面, 用户可通过数据集名称、数据集产生的时间地点、数据集提供者、数据集所属学科、关键词等进行查询, 从而方便快捷的找到需要的数据。其中, 数据集基本信息和数据集出处信息仅供浏览, 数据集实体信息只要注册为本系统的合法用户, 便可下载到本地机器上作进一步的加工处理。

(3) 系统管理。系统管理包含 2 项任务。一是用户管理。对一个数据库系统而言, 不同的用户将对数据库施加不同的操作, 依此可将用户分为不同的类型, 如系统管理员、数据录入员、普通用户等, 这是一个数据库系统正常运行所必需的。当一个用户注册到本系统后, 数据库系统要能够对注册用户进行管理, 也就是给定用户权限, 使之能够具有合理、合法的身份在允许的范围内操作数据库。二是数据集管理。一个数据集从整编、录入、校对、发布要经过若干的加工处理环节, 每一个环节都有质量保证问题, 为了很好的解决数据质量问题, 系统采用了授权和审核机制。所谓授权, 就是将操作数据集的一定权限赋予特定的用户, 如将数据集的发布权授予数据集所有者。所谓审核, 就是数据集所有者对数据集做最后的质量认定, 然后发布数据集的过程。

1.3　数据库结构的设计

数据库的结构取决于数据库所要管理的对象(或称为实体)及其各对象之间的关系, 对科技文献数据摘录数据库来讲, 其包含 3 个基本对象, 即数据集基本信息、数据集出处信息和数据实体基本信息, 数据集出处根据数据的来源不同分为 7 个子对象。为了有效的管理这 3 个基本对象, 需要构建必要的辅助对象, 这些辅助对象是: 用户基本信息、权限信息、学科代码信息、数据实体类型代码信息和数据实体下载信息。综合分析每个数据库对象所包含的数据项目及其各对象之间的关系, 即可构造出科技文献数据摘录数据库库结构设计的实体- 关系模型, 如图 1 所示。

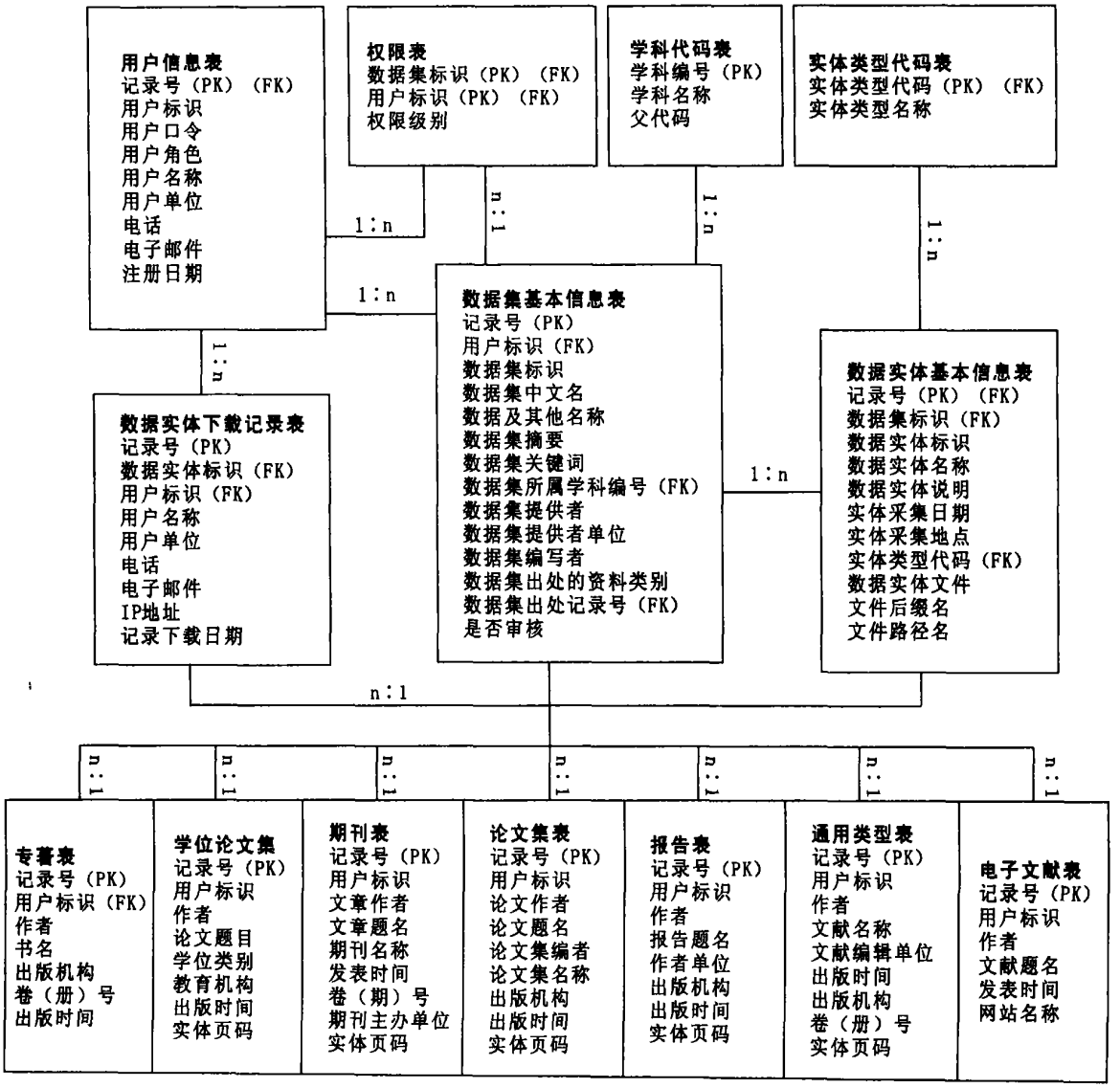

图 1　数据库实体(对象)—关系模型

图 1 中的每一个图框对应数据库中的一个基本数据表, 图框内分别为数据表名称和数据表的字段内容。图框之间的连线和标注表示表与表之间逻辑关系。根据图 1 所表示的实体- 关系模型, 便可进行数据库的开发。

1.4　数据库的安全设计与用户管理

数据库的安全性是数据库运行的基础, 也是保证数据质量的前提, 因而, 也是数据库设计必须考虑的环节。本数据库的安全性采用 Oracle 数据库系统提供的特权- 角色- 用户管理机制, 制定合理的用户管理策略。根据用户对数据库的可操作度将用户分为 4 类, 即系统管理员、数据录入员、授权用户和匿名用户。

(1) 系统管理员。系统管理员拥有对数据库系统进行维护和管理的最高权限。这些权限包括用户管理、代码表管理、数据集的修改和授权、数据集查询功能。

(2) 数据录入员。录入员由管理员授权。他可以添加、修改、删除和查找数据集。对于特定的录入员只可以修改和删除由本用户录入的数据集并且一个数据集只可以有一个录入员。这是保证数据一致性所必须的。

(3) 授权用户。用户注册成功以后即成为授权用户, 授权用户拥有查询数据集的元数据和下载已经审核过的数据集实体的权限。授权用户经过管理员授权可以升级为数据集所有者。数据集所有者拥有审核数据集的权限。从数据集的角度来讲, 数据集所有者对数据质量负全责, 并通过审核过程对数据集的质量做最后认定, 数据集一旦经过审核, 则意味着已经在 Internet 平台上发布了。

(4) 匿名用户。匿名登陆本数据库系统的用户, 只可以浏览和查询已审核过的数据集的元数据, 但不能下载数据集实体信息。

以上用户管理策略为数据安全提供了保障, 但同时也增加了数据库管理的复杂性。因此, 在进行软件的设计开发过程中要明确界定各模块的功能及其之间的关联关系, 以保证不同身份的用户登录数据库后能进入"个性化"的操作空间, 而不被无关的信息干扰。例如, 匿名用户是本数据库权限最低的一类用户, 他只可以浏览和查询已审核过的数据集的元数据, 所以, 像数据录入员所具有的添加、修改、删除数据集等权限, 系统管理员所具有的用户管理权限、数据字典管理权限等等对匿名用户就设计为不可见。这种以用户类别为主线组织数据库功能的逻辑关系, 能够使各类用户都有适合自己需要的界面和操作通道, 对数据库的建设、管理和应用都有好处。

2　数据库的应用

2.1　如何成为一个合法用户

本数据库是基于网络开发的数据库系统, 除匿名用户外的其他用户如果要使用数据库都必须经过注册, 通过有效的用户名和密码访问数据库。对于广大的用户而言, 注册是您合法使用数据库的惟一途径, 否则, 您将不能得到数据库系统提供的完全服务。也就是说, 通过匿名用户登录, 只能浏览数据集的元数据, 而经注册的合法用户不仅能浏览数据集的元数据, 而且能够下载数据实体, 恰好这是用户最需要的。另外, 本系统还有一个统计数据实体下载记录的功能, 通过分析这个统计记录, 可以帮助数据库建设者了解用户的数据需求、用户来源等信息, 以便使日后的数据库扩展更有针对性, 从而促进数据库的发展和应用。

2.2　如何建立一个数据集

本数据库的建设属于一个开放的系统, 也就是说, 如果数据库管理员赋予您合法的用户身份, 则可向数据库中添加数据集, 使您成为数据库的建设者之一。构建这种开放数据库系统的基本考虑是希望通过本数据库的建设, 搭建一个面向用户的数据管理平台, 为用户提供使用别人的数据和管理(或者发布) 自己数据的网络数据库环境, 进而促进数据共享。

向数据库中添加一个数据集可按照图 2 给出的操作流程进行, 依此用户即可将自己的数据添加到数据库中, 并且按照一定的数据共享管理政策提供数据服务。

关于数据的查询, 本系统提供了根据数据集名称、关键词、数据集提供者、数据集产生时间、数据集产生地点等检索条件, 对广大用户而言只要在相关的检索条件栏内填写自己的检索条件, 便可得到已经存在于数据库中的自己所要查找的数据。

有关数据库系统使用的详细说明, 请参考 http: //www. loess. scdb. cn/ 科技文献数据摘录数据库用户使用手册

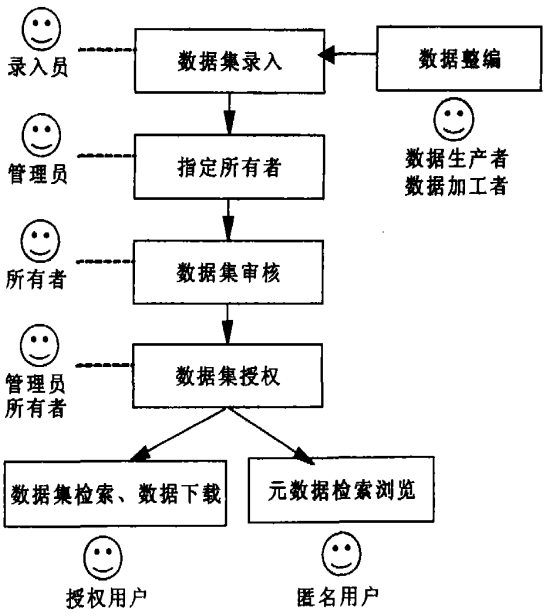

图 2　建立数据集流程

3　小　结

科技文献数据摘录数据库是基于对各类公开发表的科技文献中的科学数据特性的认识, 按照数据主题分类(关键词)、学科分类、数据产生地点、数据产生时间、数据生产者等数据的主要属性将分散在各类科技文献中的科学数据有机的组织起来, 为今后的数据参考提供查询服务的数据管理系统。它和其他的科技文献数据库的显著区别在于, 本数据库是以文献中的数据表为建设和管理的主体, 而且在数据集的整编过程中, 以黄土高原生态环境研究方面的科学数据为主要选择对象。所以, 从数据内容上讲, 本数据库是一个汇集黄土高原生态环境研究的专业数据库。从数据质量方面来讲, 所选数据都是出自公开出版物, 数据经过了作者的筛选、整理、归纳, 具有较高的科学价值。还有, 在数据集的选择中偏重生态因子、生态过程的试验、监测数据收编, 而这类数据更具有基础性、客观性、历史性的特点。因此, 对于时空跨度较大的生态研究来说, 本数据库无疑是一个具有历史价值的, 专业特色突出的科学研究基础数据库。

另一方面, 本数据库基于 B/S 体系结构建立, 并且采用了合理的数据安全和用户管理策略, 为广大用户提供了一个网络环境下查询数据和自主管理数据的数据库平台。一定程度上, 这是对正在成为发展趋势的 E- science[3] 科研环境的建立所做出的一些探索。

参考文献:

[1]　黄鼎成, 郭增艳. 科学数据共享管理研究[M] . 北京: 中国科学技术出版社, 2002. 92- 93.
[2]　孙九林. 科学数据资源与共享[J] . 中国基础科学研究, 2003, (1) : 30- 31.