

# 黄土高原植被资源数据仓库系统的研制\*

张安邦<sup>1</sup>,上官周平<sup>1</sup>,刘晓东<sup>2</sup>,赵军军<sup>2</sup>

(中国科学院水土保持研究所,陕西杨陵 712100) (西安交通大学,陕西西安 710049)

**摘要:**在总结黄土高原生态环境背景与植被营造方面的研究成果与经验的基础上,研制开发了一个由多专题、多种形式数据组成的黄土高原植被资源数据仓库系统。本系统采用基于 Internet 的分布式数据仓库模式,建立了包括黄土高原环境背景、植被种质资源、植被营造技术、植被与水保措施减沙效益等内容的植被资源数据仓库,同时运用数据挖掘技术等对数据仓库中数据进行加工,为黄土高原实现山川秀美提供宏观决策层面、技术支持层面、普通用户层面上的科技信息支持。

**关键词:** 植被资源; 数据仓库; 黄土高原

[中图分类号] S126

[文献标识码] B

[文章编号] 1007-6581(2003)03-0018-04

## 1 引言

黄土高原生态环境极为脆弱,是世界上水土流失最严重的区域之一。为了有效地改善该地区的生态环境,需集中各方面专家的知识 and 掌握区域生态环境资源的本底资料以及以往的研究成果,这些宝贵的数据多分散保存于各处,只有把这些库存的“死”数据变成可以流通的“活”数据,才能利用这些数据为西部生态环境建设服务,使黄土高原生态环境建设建立在科学基础之上。因此,我们针对黄土高原严重的水土流失和脆弱的生态环境,服务于国家实施退耕还林、天然林保护等生态工程,对黄土高原生态环境建设和植被重建的环境资源数据进行整理与分析,开展黄土高原植被资源数据仓库的建立与共享,对于退耕还林还草,防治水土流失,促进该地区生态环境整治,再造一个山川秀美的黄土高原具有积极的指导意义。

## 2 系统总体设计

### 2.1 设计原则

在系统结构设计与研制开发过程中,始终围绕如下5个原则:

(1)实用性:本项目对数据仓库设计的基本原则是在保证功效的前提下,最大可能地增强系统的实用性,用户界面友好、操作简便。

(2)规范性:本项目使用的有关土壤资源、水资源、气候资源等数据库模式和数据规范应是通用的。

(3)支持多平台:本项目实现的数据仓库允许使用多种数据库管理系统,可以在多种操作系统和网络平台上运行。

(4)稳定与安全原则:鲁棒性强、安全可靠。

(5)可扩展性:本项目实现的数据仓库要求结构合理,对环境具有较强的适应能力,便于系统功能维护和扩充。

### 2.2 系统结构

黄土高原植被资源数据仓库主要由网站服务模块、数据库查询模块、决策支持系统模块和系统维护模

块四大部分组成。

**网站服务模块:**由新闻动态、多媒体展示、区域综合治理、学术论坛、科教单位等栏目组成。

**数据库查询模块:**由种质资源数据库、植被分布数据库、林业生产数据库、气象资源数据库、土壤资源数据库、水资源数据库、耕地资源数据库、造林技术数据库、区域粮食数据库、各县粮食数据库、降雨资源数据库、区域侵蚀产沙分布数据库、科研成果数据库、植物生态文献数据库、植物生态专家数据库、科研单位数据库等16个数据库组成。

**决策支持系统模块:**由植被减水减沙效益评价决策支持系统、水保措施减水减沙效益评价决策支持系统等组成。

**维护模块:**维护模块为系统管理人员设计,便于系统内容的动态更新。系统对用户进行权限识别、判断,只有具有管理员身份的用户才能进入维护模块对系统进行维护管理。

### 2.3 系统的主要功能

(1)黄土高原植被资源数据使用和管理计算机化:黄土高原植被尤其是样本数据由于有了数据库系统,将书本资料转化为电子信息,有关的存贮、修正、查询以及统计分析等等取代了原来手工操作,具有方便、安全且快速灵活诸多优点,这是最为主要的功能之一,也是建立此数据库的目的。

(2)植被营造的环境背景查询:利用各种查询方式从不同角度获取植被营造的土壤、气象、水资源等知识,可查到有关土壤侵蚀、耕地现状、粮食生产现状、林业生产现状等。

(3)植被种质资源信息查询:这部分内容采用分类查询、综合查询、地图查询、区位查询、拉丁文查询等方式,可以查询到植被种质资源的特征、抗逆性、适应范围、分布区域、栽种技术措施等信息。

(4)植被营造技术查询:这部分内容列举了黄土高原退耕还林过程中的主要造林技术信息,查询方法简

明扼要,可以查询到具体营造技术的详细信息,为生产实践作指导。

⑥)决策服务功能: 用户可以在 Web 页上根据当地的资源现状,实现当前数据的录入、查询、修改,并直接启动决策支持系统和推理机,对事实和条件进行推理判断,得出有明确针对性的植被减沙效益、水保措施减水减沙效益,从而指导实际。

⑦)相关资料查询: 由森防动态、林业新闻、疫情通报、森林法规、药剂推荐、病虫测报、学术动态、植物学通报等几部分组成,几乎涉及到有关病虫害的所有知识,具有很高的科技含量和使用价值。

⑧)其他内容: 专家简介栏介绍森林植被生态方面专家的个人简历、科研成果和研究动向等;联系我们栏为用户提供一个留言簿,让其对网站内容提出意见和发表看法;讨论园地栏就学术界某个问题发表自己的看法。

在本站注册为会员的用户可以在论坛中可以发表文章,谈一些自己的看法或发布一些有用的信息,也可以查看其他会员的文章,专业化知识智能搜索引擎可以对站内和站外网上资源进行搜索。

⑨)数据库的查询与打印功能: 可以对各种数据库进行查询,可以查询某地区某时间段或某时间的植被资源情况,也可以了解某一类数据的总体情况,并且可以将这些查询出的数据进行打印处理。

### 3 系统实现策略

#### 3.1 系统研制的技术思路

在继承以往黄土高原植被与生态环境研究成果基础上,围绕关键科学技术问题的解决,从植被演变、种质资源、营造抚育技术、生境适应性评价、水土资源等多角度入手,以现代信息技术为支撑,沿时间序列、空间结构及时空耦合界面全方位展开研究。

采用多学科、跨学科的研究方法,在系统工程的理论框架下,将林学、植物学、植被生态学、地貌学、水文学、土壤物理学、生态学等科学方法,与实验植物学、生态工程学、信息技术等工程科学方法,和经济学、管理学方法有机地结合起来,实现不同时空尺度研究的有机集成,最终建成一个跨部门、分布式、可运行的黄土高原植被资源数据仓库共享网络系统。

(1)搜集、汇总与黄土高原生态环境演变和植被重建有关的土壤、气象、水资源、人口、农业等数据、资料、图象、图形和数学模型等信息,开展信息组织、存储、维护等分析研究,建立相应数据库。如对黄土高原地区资源树种信息数据库(树种名称、生态习性、生长习性、用途等)、自然保护区数据库(名称、位置、面积、生物种类、分布、数量等)、生态林业工程生态环境监测信息数

据库(树种组成、林分类型、径流量、小气候、土壤特性、化学特性等)和森林资源数据库(树种组成、林分类型、径级、面积、蓄积量、土地类型等)的数据结构进行分析,研究森林资源共享信息元数据的内容、格式和抽取方法。

(2)完成数据仓库的概念设计、模式设计,研究、挖掘、建立植被重建环境资源共享信息元数据的关系模式,并对关系模式进行分解和规范化,形成植被资源信息数据模式及规范。

(3)按照植被重建环境资源数据仓库标准,组建分布式数据仓库网络,尽可能将已有植被重建环境资源数据转换/包容到该网络之中。

(4)根据植被重建环境资源数据仓库的安全性、完整性、一致性、鲁棒性要求,研制该仓库共享信息管理软件。

(5)研制数据分析应用模型和数据挖掘、知识处理系统,对现有的植被的降水截留分析、数据查询分析等模型及决策支持系统进行筛选、改进,建立起相应的数学模型库、经验模型库和知识库。

(6)研制专业型知识智能搜索引擎,对植被重建环境资源数据仓库的数据以及模型库、知识库中的模型、知识进行深层次加工。

#### 3.2 数据库的结构

黄土高原植被资源数据仓库系统的数据库查询模块中共有 16 个数据库子系统,每个子系统中又对应着若干个小的数据库,所以,在整个黄土高原植被资源数据仓库系统中数据库的构建与运用成为系统中非常重要、必不可少的一部分。下面仅以黄土高原植被资源种质资源数据库的构建为例说明该系统中各数据库的构建。

##### 3.2.1 数据库的名称

该数据库的构建中涉及到 3 个子数据库,它们的名称分别为:黄土高原植被种质资源数据库[idioplasm]、黄土高原植被种质资源科名数据库[section]、黄土高原植被种质资源属名数据库[category]。

##### 3.2.2 数据的定义

◆ 黄土高原植被种质资源数据库[idioplasm]

resid	资源编号	varchar[17]	key
type	类别	varchar(1)	1-乔木、2-灌木、3-草本
sectionname	科名	varchar(3)	
categoryname	属名	varchar(3)	
resname	名称	varchar(10)	
formalname	学名	varchar(50)	
alias	别名	varchar(60)	
distributing	地理分布	varchar(400)	
feature	生物及生态学特性	varchar(1000)	

technique	造林技术	varchar(2000)
value	经济价值与实际意义	varchar(1000)
◆黄土高原植被种质资源科名数据库[ section ]		
id	科名编号	int key
sectionname	科名名称	varchar(20)
◆黄土高原植被种质资源属名数据库[ category ]		
id	属名编号	int key
categoryname	属名名称	varchar(20)
sectionid	所属科名编号	int

### 3.2.3 数据的存放

系统设计了一个系统维护模块,当需要往数据库中存放数据时,用户可在登陆时进行用户级别、权限检验。对于非管理员的用户,系统自动将页面导向一个警告页面,对于管理员,可进入维护模块直接在所选数据库对应的表格中添加数据。

### 3.2.4 数据的查询

在数据库系统的查询模块设计中,数据查询条件不是固定的。例如:在黄土高原植被种质资源数据库查询模块中,用户可以按类别、科名、属名、名称、学名、别名、地理分布、生物及生态学特性、造林技术、经济价值与实际意义这几项的任意组合进行查询,而且查询时不需要完全精确匹配。例如,某种质的生物及生态学特性为“喜光,耐寒,耐干冷气,在极端最低温度-48(C的严寒地区也可正常生长”,在生物及生态学特性栏中输入“喜温”或者“耐寒”均可查出。该模块还支持多关键字查询,在查询时将多个关键字用“、”隔开即可。例如,在生物及生态学特性栏中输入“喜温、耐寒”,表示要查询生物及生态学特性和“喜温,耐寒”均有关的种质资源的详细资料。在查询语句的选择中,我们使用的是动态SQL查询语句来实现查询。例如:查询生物及生态学特性为“喜温,耐寒”,类别为“乔木”的种质资源,用到了select查询语句。Select \* from idioplasm where feature like '%喜温%' and feature like '%耐寒%' and type like '%乔木%' by id”(注:idioplasm为种质资源表,feature为生物及生态学特性字段,type为类别字段)。

## 3.3 开发环境

### 3.3.1 运行环境

ASP程序是运行在服务器端的应用程序,硬件环境要求最好是专用的服务器,硬盘空间最少需要100M。服务器端的软件环境:Windows 2000 Server, SQL SERVER 2000,用户端的软件环境只需要一般的浏览器。

### 3.3.2 服务器端脚本选择

编程语言为ASP+, Visual C++。ASP(Active Server Page)作为主要的开发语言。ASP具有开放性、

易操作性、程序与外观分离、非手工编译、浏览器独立性等优点[9]。

### 3.3.3 开发工具的选择

选择Microsoft公司推出的Visual Inter dev6.0。Visual Inter dev6.0是Microsoft公司推出的一种新型Web应用程序开发工具,它是快速设计、建立和调试以数据库驱动的WEB应用程序的基于团队的开发环境。

## 4 结束语

黄土高原植被资源数据仓库主要根据国家科技攻关的科研成果,并总结我们以往在黄土高原生态环境变迁与植被重建方面取得的研究成果与经验,运用数据库技术、人工智能原理、数据挖掘技术等建立的一个基于Internet的分布式数据仓库系统。由于近年来黄土高原退耕还林还草工程的顺利实施,黄土高原森林植被资源现状有明显的改善,我们将对系统中的数据进行不断的更新和完善,为黄土高原实现山川秀美提供宏观决策以及为普通用户提供远程信息技术服务。

## 参考文献:

- [1] 上官周平. 黄土高原地区水分环境演变与退化生态系统恢复[J]. 中国历史地理论丛, 2001:218~221
  - [2] 上官周平, 李秧秧. 控制水土流失, 整治黄土高原生态环境[J]. 科学对社会的影响, 2001, (1):16~19
  - [3] 中国科学院黄土高原综合科学考察队. 黄土高原地区自然环境及其演变[M]. 北京: 科学出版社, 1991.
  - [4] 吴钦孝, 杨文治主编. 黄土高原植被建设与持续发展[M]. 北京: 科学出版社, 1998.
  - [5] 杨宝祝, 赵春江等. Web数据库应用实现技术[J]. 高技术通讯, 2000, 10:51~53.
  - [6] 清汉计算机工作室. ASP开发实例[M]. 北京: 机械工业出版社, 2000
  - [7] 蒋定生主编. 黄土高原水土流失与治理模式[M]. 北京: 中国水利水电出版社, 1997.
- \* 教育部中央财政专项和中国科学院自然资源与环境知识创新工程重要方向资助项目(KZCX02-411-04)资助项目

## The Design and Implementation of the Data Warehouse on Loess Plateau Vegetation Resource

ZHANG An-bang SHANGGUAN Zhou-ping  
(Institute of Soil and Water Conservation, Chinese Academy of Sciences, Yangling, Shaanxi, 712100 China)

LIU Xiao-dong ZHAO Jun-jun  
(Xi'an Jiaotong University, Xi'an, Shaanxi, 710049 China)

**Abstract:** The paper summarizes the research results and experiences of ecology environment and vegetation rebuilding in the Loess Plateau, establishing a composite database

composed of many special topics and forms. Based on the internet, we establish a vegetation resources data warehouse including Loess plateau vegetation distribution, composition and function, vegetation resources bearing capacity, water and fertilizer transferring processes in vegetation rebuilding and others, at the same time the data in the database are also

more handled and used at the help of artificial intelligence theory, data mining technique and other technique, for realizing hills and plains beautiful offer scientific message support for macroscopical policies, technology supports and common users.

**Keywords:** Vegetation resource ;Data warehouse ;Loess plateau

## 如何提高扫描仪的文字识别率

楚杰

扫描仪的一个重要功能,它使印刷体文字的录入最终摆脱了键盘。但是在OCR识别的过程中,往往会出现许多识别错误的情况,这到底是什么原因造成的呢?我们又应该采取哪些措施来提高OCR识别率,以减少人工修改的麻烦呢?下面是笔者在使用过程中积累的几点经验,供各位参考。

### 1. 选用专业的OCR软件

购买扫描仪时,附带的扫描软件中普遍都提供OCR识别软件,但是其识别率很难令人满意,所以一般应当购买和使用专业的OCR识别软件。

### 2. 在识别前一定要作版面分析和倾斜校正。

### 3. 分割识别混排文稿

简繁混排和中英文混排的文稿在识别时往往会出问题,如果文稿中简繁/中英文是块状分布(即这几段为A文,另外几段为B文),则可以采用图形处理软件把块与块分开成独立的文件,然后分别打开,选择相应的语言,分别识别。

### 4. 充分使用识别软件的用户自学习功能

OCR识别软件内一般都含有用户自学习功能,如果个别字在扫描比较清楚的情况下,软件识别得不正确,就可以把这个字加到用户自定义库中,这样可以逐渐地提高识别的正确性。

### 5. 尽量选用质量较高的图像进行识别

要提高OCR识别能力,在很大程度上取决于

识别所用的图像质量,在尽可能的情况下应选择质量较高的图像进行OCR识别。要想获得较高的扫描识别率,用户应当不断进行尝试,找出一个适合OCR识别软件的扫描分辨率及扫描亮度以供参考。

### 6. 选择适当的扫描分辨率

选择合适的扫描分辨率:不求最“高”,只求最“佳”。如果分辨率太高,扫描的时间将会大大增加,图形所需的空间也会成倍地增长;如果分辨率太低,给OCR软件的信息量不足,识别率就不会太高。根据经验,普通五号印刷体采用250~300dpi比较合适;若字号比较大(四号以上),用150~200dpi就足够了;如果是六号或七号字,就要考虑使用400~600dpi了。当然,这也不是绝对的,如果印刷字迹比较模糊,或者笔划较多的识别文本,应适当提高扫描的分辨率。

### 7. 设置合理的扫描亮度

有一些印刷品会因纸质发黄或是有较多的油墨点而影响识别质量,因此在扫描时适当地调整亮度,使扫描文件黑白分明,有利于软件的识别。扫描亮度的设定以观察扫描后的图像中汉字的笔画较细但又不断开为原则。对于扫描后图像存在黑点或黑斑时,应该适当减小亮度;如果当扫描后图像中汉字轮廓严重残缺时,则可以适当增加亮度。