

论关系型园林植物数据库建立与检索

吴福明

(中南林业科技大学 环境艺术设计学院 湖南 长沙 410004)

摘要:通过对现有园林植物数据库的分析,指出了园林植物数据库的数据结构特点及其特有的数据检索方式,探讨了建立关系型数据库表的基本方法,提出了基于植物的科学(自然)分类法的“通用数据库接口”方案来解决各种数据库之间的数据共享问题。

关键词: 园林植物数据库; 植物检索表

中图分类号: TU 986.3 **文献标识码:** A **文章编号:** 1001-0009(2010)12-0231-02

随着电子计算机技术的发展,各种计算机软硬件生产渐趋成熟,价格越来越便宜,计算机得以进入了千家万户。人民的生活水平得到了极大的提高,对环境的要求也越来越高,园林植物成了寻常百姓家的不可或缺的重要装饰和摆设。由于园林植物也像计算机一样进入了千家万户,人们对园林植物的认识得到了很大的提高。计算机、因特网、园林植物成了百姓家耳熟能详的普通事物。园林植物的识别不再是少数科学家的事,越来越大众化、市场化,因此建立一个高效的园林植物检索数据库将具有美好的前景和广阔的市场空间。

经过几十年的发展,计算机数据库技术已日趋成熟,自从因特网技术兴起之前的单机数据库到现在的分布式数据库技术,从简单的本机单表数据库到关系型数据库再到庞大的网络分布式数据库,无数的植物学专家和工程技术人员开发了为数众多的各种各样的、不同用途的数据库系统。比较成功的有《中国数字植物标本馆》、《中国园林网植物库》、《PPBC 中国植物图像库》等基于 WEB 网页的植物数据库查询系统。植物学者们也在不断地探索与实践中,如马莎、赵大志等的攀枝花苏铁特色数据库建设^[1],何绍顺的植物电子图像标本与造林树种档案数据库构建技术^[2]。由于这些植物查询系统的独立性和互不兼容,彼此之间不能有一个通用的、标准的数据格式,因此相对来讲这些系统还是一个独立的相对封闭的查询系统,其应用受到了其自身的限制而彼此之间不能互通有无,因此采用通用“数据接口”来解决数据通信与数据共享成为当务之急。

1 园林植物数据库数据分析

1.1 基于“种”的关键词(KEY)

植物的种 (Species) 应该是其最小能够检索的单位^[3]。在“种”下分为亚种 (Subspecies)、变种 (Varietas) 和变型 (Forma)。目前在“种”上存在的最大问题是“别

名”问题。使用拉丁名可以有效地避免这个问题。但是广大人民群众大多不是植物方面的专家,对于这些晦涩的拉丁名人们只能敬而远之。这不利于数据库系统的普及与商业化。另一种生成 KEY 的方法是编码法。比较典型的是中国科学院植物研究所和中国标准化与信息分类编码研究所于 1994 年建立的中国植物分类与代码国家标准 (GB/T 14467-93)。该标准采用了层次编码法,每一完整代码都由 10 或 12 位阿拉伯数字组成。同样这些无法从字面上看出任何意义的编码也像看拉丁文一样令人生厌,此标准并没有大规模的实际应用。汉字才是中国人喜闻乐见的植物名称,植物的中文名是广大劳动人民在长期的生产实践中总结出来、形象化的语言。因此园林植物中文名的唯一性工作是建立园林植物数据库首先要解决的问题。但是由于植物分布地域的广泛性,不同地区的人们对同一种植物的名称又有所不同,因此出现了“别名”。如果使用特定的别名在某些网络植物数据库中是不能检索出来的。这称为“别名局限性”。植物的中文名称应该是这些别名中具有唯一性的任一名称。有必要对植物拉丁名的数据结构进行分析以应用于对植物中文名进行重新命名。

1.2 数据冗余

植物的检索一般是基于植物形态的。特别是专用于植物形态识别的专用数据库,主要检索的关键词是植物的形态特征描述文本。但是有许多植物的形态描述文本是相同的或大部分相同。如同科同属的植物。如果以“种”为基本单位建立数据库则会造成大量的数据重复。这些重复的数据加大了数据库的体积,从而降低了检索的效率和速度,限制了植物数据库的大小。实际表现是随着数据库体积的增长而检索性能直线下降。

一种有效的方法是以“科”、“属”、“种”三级单位组织数据库可以有效减少数据冗余,这在现有的一些植物方面的书籍中得到了体现,如《中国树木志》(陈有民著)等大批书籍都是采用此种方法。“科”以上的数据单独以关系表的形式构建。

作者简介: 吴福明(1970-),男,本科,研究方向为环境艺术设计。
收稿日期: 2010-03-31

1.3 树形数据结构

植物的分类图的形状由于其进化历程的渐进性基本上是树形的。任何一种植物都能在这棵“树”上找到属于它的一个节点以及它与其它植物共用的上级节点(界、门、纲、目、科、属等节点)。所有的节点都共同具有一个根节点。根节点是这棵树的第 0 层, 以下各层依次应为: “门”为 1 级、“纲”为 2 级、“目”为 3 级、“科”为 4 级、“属”为 5 级、“种”为 6 级共 7 级节点及其各级所具有的“亚”级。任何节点所处的“级”数又称为“节点深度(NODE LEVEL)”。

由此, 植物数据库基本上具有天然的“层次”结构。由于树形数据结构的数据存取算法较为复杂, 基本上处理“上下级”的方法为“关系表法”。

1.4 多分类系统的网络数据结构

由于一个数据库系统不可能采用单纯的“科学(自然)分类法”来组织整个数据库中的所有数据, 科学(自然)分类法中的同级“节点”之间可能发生横向“关系”, 不同级“节点”之间可能发生斜向关系。这就打乱了严谨的“科学分类法”中的节点关系, 从而形成了纵横交错的“网络表关系”。由此, 采用多分类系统的数据库适合采用“网络表法”。网络法的数据库结构更为复杂, 数据库设计难度更大。

2 园林植物数据库的特点

园林植物数据库降低了具有以上的通用植物数据库的特点外, 从专业角度出发, 为了满足专业人士及广大人民群众的实际需要, 还必须具备以下特点。

2.1 基于检索表的二叉(树形)检索方法

园林植物数据库的主要功能是为了获得植物的观赏、栽培、管理、用途、识别等为园林设计、植物配置等专业信息服务的数据库, 是通用植物数据库的一个子集。根据其用途和数据特点可分为植物识别与分类数据库、植物图像数据库等。例如植物识别数据库的表构建, 其字段划分主要依据是识别的手段需要。园林专业人士主要使用传统的植物检索表来识别植物, 传统地检索表有平行检索表与等距检索表 2 种。对植物检索表的数据结构进行分析有助于园林植物识别数据库的表设计和字段设计。这种二叉式的检索方法(数据类型是布尔型):

1. 植物体无花、无种子, 以孢子繁殖。
2. 植物体有茎、叶分化或扁平的叶状体, 无真根和维管束 苔藓植物门

2. 植物体既有茎、叶分化, 也有真根和维管束 蕨类植物门
1. 植物体有花 以种子繁殖。
- 3 胚珠裸露, 不包于子房内 裸子植物门
- 3 胚珠包于子房内 被子植物门

从以上定距检索表中可以看出裸子植物与被子植物的区别就是一个布尔型数据: 包于子房内和不包于子房内(无子房)。每个布尔型数据下有 2 个分枝, 这 2 个分支是二岐型的, 也有多岐型的数据(略)。

2.2 基于图像等的感性认识数据检索

图像的自动识别系统是基于大众感知的识别系统。植物识别不仅仅通过形态识别, 还可以通过植物的树形、颜色、质地、纹理等其它“感性”特征, 这些特征是传统检索表中所没有的。这些数据特征往往是“感性”的, 甚至不是基于文字的(无法用文字来描述), 这些数据不利于数据库的检索。园林植物数据库往往是从其观赏特性(其中大部分为“感性”数据)出发的数据库系统, 这些“感性”数据在计算机中主要是通过图像来存贮的。一个长期从事园林花卉生产的工人往往不能很好的使用植物检索表, 但是却能通过“感觉”来识别大量的植物, 必须承认“感性认识”的存在。建立植物颜色数据库、植物表面纹理数据库(如树皮纹理、叶脉)、叶形特征数据库等以提供辅助检索手段^[4]。

综上所述, 园林植物数据库的建立必须是以“种”为基本检索单位的具有复杂的数据表结构的“科学”与“抽象”的结合。其检索也是建立在“科学(自然)分类法”的基础之上的灵活的检索系统。为了确认任何物种在数据库中的唯一身份, 使数据库的完整性和唯一性得到实现, 方便各种植物系统之间的互通和联合, 可以充分利用植物的“科学(自然)分类法”建立各种数据库之间的统一的检索“接口(INTERFACE)”, 通过这个“接口”访问绝大多数植物数据库系统为数据共享的基础。

参考文献

- [1] 马莎 赵大志, 陈吕容, 等. 攀枝花苏铁特色数据库建设[J]. 图书馆 2007(5): 107-110.
- [2] 何绍顺. 植物电子图像标本与造林树种档案数据库构建技术[J]. 林业调查规划 2006, 31(1): 30-33.
- [3] 王彦勋 王利民, 孙宏凯, 等. 二叉树数据结构在植物检索中的应用[J]. 安徽农业科学 2008 36(18): 7872-7873.
- [4] 林大辉 宁正元. 基于图像的树种分类方法的研究[J]. 福建电脑 2008(1): 34-65.

Retrieval of Garden Plants on the Relational Databases

WU Fur-ming

(School of Environment and Art Design, Central South of University and Technology, Changsha Hunan 410004)

Abstract: Through the analysis of the existing garden plants database, and pointed out that the garden plant database, the data structure and the unique characteristics of data retrieval methods to explore the establishment of relational database tables of the basic method was proposed based on plant science (natural) classification of “universal database interface” option to solve the data sharing between databases.

Key words: garden plant database; plant keys