

# 活植物信息数据库中的植物编码系统

顾 烟 高秀梅 凌萍萍 贺善安

(江苏省植物研究所, 南京 210014)

**摘要** 本文概述了南京中山植物园活植物信息数据库中的植物编码系统, 重点介绍了植物编码的生成方法。该系统通过将关系数据库 ORACLE 与 Microsoft C 所提供的完备的计算模型与关系查询语言的强功能管理手段结合起来, 弥补了现行商用数据库的不足, 获得了较高的时空效率。

**关键词** 关系数据库; 植物编码系统; 编码生成方法

**The plant systematic code system of the database for living collections** Gu Yin, Gao Xiu-Mei, Lin Ping-Ping and He Shan-An (Jiangsu Institute of Botany, Nanjing 210014), *J. Plant Resour. & Environ.* 1993, 2(2): 1~5

The plant systematic code system of the database for living collections in Nanjing Botanic Garden Mem. Sun Yat-Sen was summarized and the coding method was presented in this paper. The code system was programmed by combining the ORACLE relational database together with the Microsoft C which provided the perfect calculating mode and the strong management technique of relational inquiry language. It prevented from the disadvantages of present commercial database and presented a higher efficiency both in time and space cost.

**Key words** relational database; plant systematic code system; coding method

本世纪80年代以来,随着人类环境意识的日趋强烈和对植物资源流失的关注,植物园被推到了保护植物物种的前沿,从而导致了全世界植物园数量上的大大增加,而计算机技术的发展又使得计算机的应用由单一的科学计算,发展到各行各业的各个方面,它在植物园活植物信息贮存与交流上的作用也日趋显著。在植物园活植物信息数据库中,植物名称是必不可少的项目,而且一般都作为关键字段在数据库中起十分重要的作用。因此,对其进行科学编码已成为开发植物园活植物信息数据库的一个制约因素。目前,已有一些单位在进行植物名称的编码工作,但由于其对科以上的等级、定名人、种下单位及品种的编码以及杂种编码重视不够,从而使得其编码的使用带有很大的局限性,而且现行的编码系统对每一植物种类的编码范围作了限制,只给将来的增补工作留下较小的余地。因此,作者等与南京航空学院计算中心合作,构建起一个植物编码系统,该系统建立起一个植物分类信息与编码的一一对应关系,用格式确定、长度一致的编码代替冗长复杂的植物学名及其在分类系统中的亲缘关系。使用时,编码系统将接收的植物分类信息转化为编码,其相应的操作转化为库中对编码的操作;对于查询后返回的编

码,可利用编码系统的逆运算机制,得到其显式的植物分类信息(如学名)。

## 植物编码系统中包含的植物分类等级

根据国际植物命名法规及国际栽培植物命名法规可知,植物名称可用BNF范式归纳为:  
植物名称 = <属名> <种加词> <定名人> { <种下单位加词> <定名人> } (品种名称) (式1)

种下单位 = <亚种> | <变种> | <变型> (式2)

《植物园植物记录国际传输格式》(‘The International Transfer Format for Botanic Garden Plant Records’,简称ITF)规定,在植物园活植物信息数据库中,对于象科名这样较高的类群的处理,可以作为附加字段,而在传输文件时删去。然而,一个足以说明植物亲缘关系的分类系统,对于植物园植物引种驯化、培育改造以及植物资源的发掘利用乃至科普教育等工作来说,都是重要的科学内容。因此,在植物园活植物信息数据库的植物编码系统中,包括科以上的各级单位是很必要的。

植物分类单位的主要等级自上而下依次是:界(regnum)、门(divisio)、纲(class)、目(ordo)、科(familia)、属(genus)、种(species)等,它们之间的关系象一棵倒立的大树,因此编码的关键便进一步转化为数据库中树型结构来表示和实现植物分类等级间的树型结构。

## 植物编码的生成和翻译

在关系数据库中,实现树型结构一般采用压缩法。但压缩法是建立在严格的关系模型基础上的。由于关系模型的表(二维)结构和应用模型的树结构之间存在着天然的差异,实现时必须将应用模型的树结构转化为关系模型的表结构,这种转化不可避免地是以时间或空间的损失为代价的,空间损失的原因在于数据模型的不一致,时效损失的原因在于其计算模型不完备<sup>[9]</sup>。由此我们给出了植物分类等级的层位码压缩方法自动实现植物编码。该方法通过ORACLE的PRO \* C产品,将Microsoft C所提供的计算模型与关系数据库查询语言的强功能管理手段结合起来,构成一个时空综合效率高的编码系统。

此系统的核心在于植物编码本身表示了各节点所处的层次及其在层次中的位置,并采用了字符串的编码方式。在编码的制定过程中,还对植物的各分类等级的数量进行了充分的统计,以确定各层次编码的位数,所以此编码系统能应付所有的应用需要。而且它的C语言与ORACLE结合设计的形式决定了此编码系统既可单独使用,又可移植到与植物有关的其他数据库中使用。

层位码的实现方法是将植物的各等级分别设置成一张表,其结构如表1。

表1 层位码中的表结构

Tab 1 List structure of level & position code

节点编码	Nodecode	节点信息	Nodeinfor
.....			.....
.....			.....

编码的生成可用公式表达为:节点编码 = 父节点编码 || 节点序码 (||为连接操作符)。其节点编码是植物编码中的一个子串,它的长度依层次的增加而增长,当到达植物的最低分类等级时其节点编码就是该植物的编码。而当植物

分类系统中所有等级的表创建完毕后,便形成了完整的植物编码系统。

在植物编码库的建立过程中,分别根据 Brotherus 系统(1924~1925)、秦仁昌系统(1978)、邢万钧系统(1975)及 Engler 系统(1964)对苔藓类植物、蕨类植物、裸子植物、被子植物(低等植物待补充)的科以上单位进行编码,属以下各单位的编码根据其插入先后自动生成,而在逻辑上则根据其编码的大小排序,这也就避免了以往的编码系统按字母次序依次排列连续编码而引起的编码扩充后的一致。

编码生成以后,复杂的植物分类信息在系统内部便有了一个统一而简洁的内部表示,使用时可通过截取植物编码来求得各节点的编码,并由此查得各节点信息,将得到的所有节点信息按序拼接起来,便可获得其完整的植物分类信息(如学名)。植物编码的翻译虽然要打开每一层次的节点表,并在多个表中查询,但这不同于横压法中必须一层层遍历完整路径的查询方法,层位码到任何层次节点编码的映射只须一次串的截取操作,同时可获得所有层次的节点编码,加上 ORACLE 一次可打开多张表,其搜索深度为1,具有较高的时效。

与此同时,考虑到在植物学名中定名人的引用在实际应用中的意义,我们舍弃了 ITF 中植物名称不包含定名人的方法,将定名人引入植物编码系统。为了避免引用定名人缩写的差异而引起的混淆,并减少冗余存贮,我们建立了一个定名人字典,将植物学名中经常重复出现的定名人信息存于字典中。

表2 纵压法中的表结构  
Tab 2 List structure of longitudinal compress technique

门	Divisio	纲	Class	目	Ordo	科	Familia
.....	.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....	.....

特别值得一提的是,在此编码系统中,我们采用纵压法将门、纲、目、科的信息压缩到一张表中。它将树结构中一条完整的路径作为表中的一行,路径上每一层次的节点作为表中的一列,纵压法中的表结构如表2。

因此,原来不同层次的信息在表中均处于同一层次,其搜索深度为1,具有很高的时效,又因科的总数不多,所以其冗余存贮数量较小,但其表结构却很容易实现。

同时,考虑到种下分类单位和品种数量很大,其高度不大于2(见式1、式2),所以采用横压法将种下分类单位和品种存于同一张表中,其结构如表3。

由此可见,表中的一行只表示树结构中的一个节点,各层次的节点均存于同一张表中,其中当前节点信息记录节点值,父节点编码记录该节点的父节点在表中的编码值,通过这个父节点编码,便可使表中一个个分离的节点信息连接起来,还原成树结构。由于横压法中一行只代表一个节点,所以它克服了纵压法中重复存贮的问题,这样虽然搜索深度增加了1(因每个节点仅“记”得其父节点,编码和名称的转换必须遍历整个路径,依次查询各节点信息,搜索高度为树高),但却减少了一张表,减少了冗余存贮,且可直接利用 ORACLE 所提供的标准实现。

表3 横压法中的表结构  
Tab 3 List structure of horizontal compress technique

当前节点编码 Nodecode	当前节点信息 Nodeinfor	父节点编码 Fanodeinfor
.....	.....	.....
.....	.....	.....

### 杂种名称的编码问题

本编码系统提供了处理杂种(包括有性杂种和无性杂种)名称的能力,其科以上单位编码的生成和翻译与上述情形相似,属以下各单位的编码生成和翻译随杂种的命名方法而不同。

### 1. 杂种公式

在处理杂种公式时,我们采用的是使用两个记录的方法,即分别输入母本和父本的学名,形成一个杂种编码。显示和打印(编码翻译)时,则通过截取杂种编码来求得各节点编码以求得各节点信息,将得到的所有节点信息按序拼接起来,便得到用连接码(× 或 +)连接母本、父本的名称(杂种公式)。如是种内杂种或同一种内种以下诸分类单位间的杂种,可根据命名法规,自动将父本的属名、种加词、定名人等略去或用缩写,如:

*Salix aurita* L. × *S. caprea* L. 或 *Salix aurita* L. × *caprea* L.

### 2. 拉丁文集合名称

因杂种被给予一个拉丁学名,所以处理拉丁文集合名称也就变得格外简单,其杂种编码的生成与翻译方法与正常植物学名相似。只要根据它是属间杂种、种间杂种还是种以下的诸分类单位间的杂种分别在不同的字段上加上杂种标记,输出时连同标记一起输出即可,如:

× *Heucherella tiarelloides*

*Salix* × *capreola* Kerner ex Andersson.

*Polypodium vulgare* subsp. × *mantaniae* Rothm. ex Schidlay

### 3. 非拉丁文集合名称

因为使用的是相当于拉丁文集合名称的现代语言,所以其编码的生成与翻译方法与拉丁文集合名称相似,只是它的杂种标记为 Hybrid, Cross, Crosses, Grex 等,如:

*Lilium Bellingham* Hybrids

*Rhododendron Fabia* Grex

以上2,3二种方法讲的都是直接杂交的例子,还有一些集合名称包括比较复杂的家系,其亲本有杂种。如 *Rhododendron Jalisco* Grex, 是下面四个种的复合杂种:R. "Dido" (*R. decorum* × *R. dichroanthum*) × R. "Lady Bessborough" (*R. campylocarpum* var. *elatum* × *R. discolor*)。对于这种情况,该编码系统提供了一种递归的方法记录和查找其母本、父本的学名,其编码库的结构与杂种公式的编码库结构相似。

## 总 结

通过对以上植物编码系统的描述可知,本编码系统具有如下优点:

1. 本系统通过将 ORACLE 的 PRO \* C 产品,与 Microsoft C 所提供的完备的计算模型与关系查询语言的强功能管理手段结合起来,弥补了现行商用数据库的不足。

2. 在植物编码系统中,总体上采用的是层位码法,但考虑到具体情况又结合采用了压缩法,使得时空效率大大提高。

3. ORACLE 数据库的变长记录格式,在植物编码系统中显著地提高了存贮空间利用效率。

4. ORACLE 数据库具有通用性强,大、中、小、微各档计算机兼容,植物编码系统的可移植性强。

5. 本系统具备处理植物的分类单位的各等级、品种及杂种的编码的能力。

由此可见,在新一代数据库系统尚未成熟时,该植物编码系统具有很强的生命力和实用价

值。

### 参 考 文 献

- 1 赵士洞译. 1984; 国际植物命名法规, 科学出版社, 北京.
- 2 黄志球, 尤晓梅, 刘永志等. 1992; ORACLE 中树型结构的实现, 第七届全国管理信息系统学术学会论文集, 101页.
- 3 顾 姻, 贺普安, 1990; 植物园植物记录计算机管理系统, 河海大学出版社, 南京.
- 4 Ruth Kvalen. 1984; *Plants and the Landscape*. 7(2): 19~20.

(责任编辑: 管晓春)

## 环境、资源与植物资源学的概念

周荣汉

(中国药科大学植物化学分类研究室, 南京 210038)

开发利用植物资源是当今的热门话题, 涉及的问题很多, 现仅就有关环境、资源及植物资源学等概念问题谈谈看法。

**环境与资源** 以人为主体的, 人类生存的周围空间及空间介质称为环境。它包括自然环境和社会环境。自然环境由许多环境要素组成, 如大地圈、岩石圈、水圈和生物圈。“资源”是泛指可供利用的天然物质资源和能量资源。

在自然界里并非所有的物质和能量都称为资源。那些对人们无用(或暂时无法利用)甚至有害的则不叫资源。如洪水、猛兽在古代显然不能叫资源; 在现时, 洪水可用来发电, 称为水电资源; 猛兽被制服, 可成为野生动物资源。但地震、雷电释放出的能量及一些毒草(如豚草)现时尚未能利用, 故不能称为资源(不能称地震资源、雷电资源、杂草资源)。所以, 人类从环境中索取资源的种类、数量、规模、范围以及利用的程度, 都取决于社会人口数量、科技水平和生活水平, 即整个人类的发展状况。社会在进步, 资源在发展, 人类对资源的认识、开发与利用(广度和深度)是无穷尽的。

**植物资源** 环境中可利用的植物称为植物资源, 既然资源可利用, 具有满足人类需要的功效, 因而是有价值的。其价值包括比较实的物质性价值及比较虚的生态环境价值。前者较易量化, 后者难以核算, 但两者都是实际存在的。所谓物质性价值不仅包括资源产品的价值(资源开发的成本和利润等), 而且应包括植物资源本身的价值。有价值的植物资源应当是有偿

占有, 并应制定一系列法规加以保护, 如森林法、草原法、野生植物保护法等, 否则植物资源将会遭到掠夺性的开发和浪费, 以致造成资源破坏、生态平衡失调和环境恶化等。环境要保护, 资源要开发, 人们需要从自然界不断获取大量的物质原料, 因而加强环境科学及资源学的研究和人才培养, 实为当务之急, 也是造福后代的大事。

**植物资源学** (Resource Science of Plants = Phytoresourceology) 植物资源学是研究资源植物的种类、数量、地理分布、时(间)空(间)变化、合理开发利用和科学管理的学科。其目的是在研究资源植物分布规律的基础上, 运用经济效益的优化技术, 合理安排资源植物的采收、加工和综合利用等, 使社会效益、经济效益和生态效益三者协调发展, 试图谋取更多的有用物质与产品, 以满足人类的需要。

植物资源学是一门综合性学科, 更确切地说是一个多学科、跨学科兼有管理学科性质的综合性学科。与经济植物学(Economic Botany)、植物化学(Phytochemistry)等学科有不同的范畴和任务。它的基础主要是生物学、生态学、生物化学、天然产物化学(Natural Product Chemistry)及工艺学等。

植物资源学的主要任务是资源调查、资源开发与综合利用、资源的经营管理以及新资源的寻找与扩大等。

(在全国植物资源开发利用学术讨论会(1992, 11. 20, 北京)上发言的一部分)