

# 英、拉、汉树木名称电子词典 TreeName 的研制

郑勇奇<sup>1</sup>, 张川红<sup>1</sup>, 郑洪涛<sup>2</sup>, 郑志华<sup>2</sup>, 李伯菁<sup>1</sup>

(1. 中国林业科学研究院林业研究所, 北京 100091; 2 北京航空航天大学, 北京 100082)

**摘要:** 英、拉、汉树木名称电子词典第 1 版(TreeName 1. 0) 具有树种的英文、拉丁文和中文名称的相互翻译查询功能。软件包含了 1. 5 万余条英、拉、汉树木名称词条, 能够进行快速有效的检索查询, 为工作提供极大的帮助。整个软件采用基于对话框模式的查询界面和基于文件系统的数据库作为整个查询系统的框架。本系统在设计中采用了比较灵活的功能模块设计, 利于软件的更新。与印刷版的各种词典相比, 电子词典系统具有无法比拟的优点, 它能够及时进行修改、补充, 使系统不断得到完善, 及时根据用户的反馈信息进行改进, 有利于软件质量的提高和功能的完善。

**关键词:** 树木名称; 电子词典; 查询; 数据库

中图分类号: Z38 文献标识码: A

树木拉丁学名在开展国际林业学术交流方面起重要的作用。由于多数树木(植物)分布范围广泛, 在很多不同地方生长, 在不同地方有不同的名称, 包括汉语名称和英语名称, 因而很容易产生混乱, 只有拉丁学名才是标准的名称, 全世界都是一个标准, 从而能够解决名称混乱的问题。我国出版过多种关于树木英、拉、汉名称的工具书<sup>[1~4]</sup>, 但各有侧重点, 所包括的树种不全面, 因此, 一个林业工作者往往需要准备多种树木名称工具书, 即使如此, 还远不能满足工作需要。当遇到生疏树种名时, 疲于查阅大量树木名称工具书, 费时费力, 效率低下。对于非树木分类的人员或非林业人员, 从工具书中查询树木名称更不是一件易事。如果没有一本全面的拉、汉、英树木名称对照目录, 势必给从事林业工作的人员在工作中带来许多不便; 另一方面, 由于树种的数目很大, 受篇幅所限, 印刷版词典很难收集齐全的树种名录。

随着计算机的普及与信息技术的迅速发展, 在查阅文献资料方面计算机的应用也越来越广泛。国内已有多种中、英对查的电子词典, 如金山词霸、网际金典<sup>[6]</sup>、东方快车<sup>[7]</sup>等软件相继开发面世, 但用于具体行业的专业词典很少, 尤其是林业专业领域的电子词典目前尚无报道, 因此, 急需研制开发相应的树木名称电子词典, 为广大林业工作者提供方便。树木名称电子词典能够使用户很轻松地实现各种树木的中、英、拉三种文字的名称对查, 大大提高工作效率, 再也不必花费精力去记忆大量的树木拉丁名和查阅许多工具书。英、拉、汉树木名称电子词典研制, 是计算机技术和信息技术在林业领域的具体应用, 也是对出版印刷树木名录传统方法的创新之举。本文向读者介绍该软件的研发原理、功能及使用方法, 并探讨软件进一步扩展的可能性。

收稿日期: 2003 05 20

基金项目: 中国林业科学研究院科技发展基金资助(2003)

作者简介: 郑勇奇(1964—), 男, 湖北天门市人, 研究员, 首席专家。

## 1 软件功能

英、拉、汉树木名称电子词典第1版(TreeName 1.0)具有树种的英文、拉丁文和中文名称之间的相互翻译查询功能。词典收集了近万条树种名称词条,查询界面简单明了,操作方便易学,并具有词库扩充、在线帮助等功能。本电子词典除具有正常印刷词典的各种树木名称查询的功能外,还能提供树种的描述信息。利用本词典,能够进行快速有效的英、拉、汉树种名称对查,并获得树种的相关信息,为工作提供极大的帮助。

目前的电子词典所具有的主要功能如下:(1)树种英文名、汉语名和拉丁文名之间的互查(包括模糊查询);(2)树种简要描述,如形态特征、自然分布、气候条件等;(3)词典的更新,用户可以建立自定义词库以添加新词;(4)热键帮助系统,支持上下文敏感的帮助,即F<sub>1</sub>功能键;(5)可扩展性强,系统预留功能扩展接口,为将来添加多媒体信息,如图片、视频、音频数据等。亦可方便地移植转变成基于网站的网络版本;(6)其它Windows操作系统下应用程序的常用功能,如常驻内存,“缩小在”任务栏上,并支持热键启动等功能。

## 2 软件设计原理和 workflow

### 2.1 总体框架

整个体系结构分为3层,最下面是数据库,中间为查询层,最上面为界面层。

用户界面层的功能是显示人机对话界面,接受用户的输入和指令并执行,返回并显示结果,提供帮助系统、热键、驻留内存等支持。查询层的功能是把用户查询和更新、添加新词、用户自定义词等的操作转换为对数据库的操作。数据库层负责对数据库的操作,包括查询、添加、修改、删除、一致性检查等。

整个软件采用基于对话框模式的查询界面和基于文件系统的数据库作为整个查询系统的框架。采用基于文件系统的数据库,使用起来方便快捷,便于高效查询。将词汇存放于数据文件中,软件在程序启动时,将数据库文件中的词条信息数据读入内存,以提高检索的效率。在查询到词条信息后,根据查询到的词条索引信息,直接去数据库中的指定位置读取整个词条的全部信息。

具体结构和查询过程如图1所示。

### 2.2 数据结构和查询方法

2.2.1 数据的存储结构 数据分文件数据和内存数据两种。文件数据由4个文件组成,其中有三个文件存储的分别是中文、英文、拉丁文所有的词条记录,这些词条在文件中按顺序排列,以供数据加载进入内存时构成有序的数组。每个词条记录结构为: {词条,索引},此结构作为文件的存储单元。代码如下:

KEYLEN 为词条的最大字符数; VALUELEN 为中文解释的最大长度。

```
struct temKeyUnit
{ unsigned int m_Address;
  char m_Key[KEYLEN]; }
```

还有一个文件存储的是中英拉中文解释索引的记录,此文件系统可以通过索引查询访问。存储代码:

```
struct ValueUnit
```

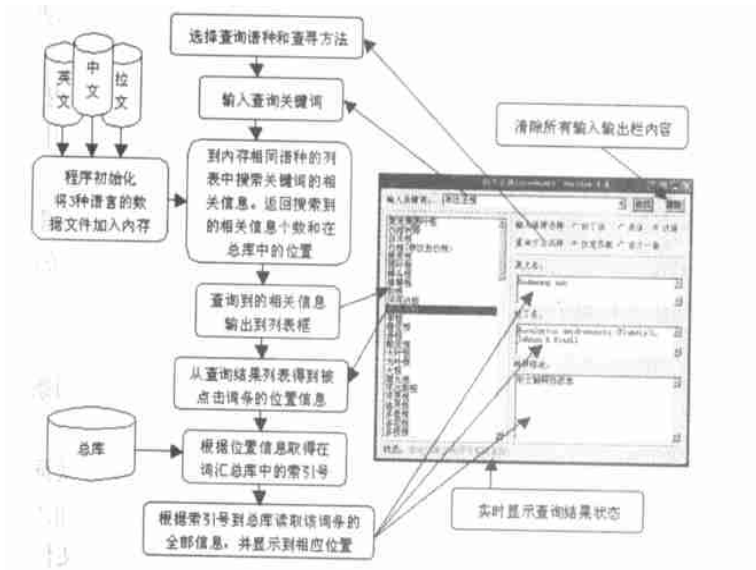


图 1 树木名称电子词典的具体结构和查询过程

```
{ unsigned int m _ Address;
char m _ EngKey[ KEYLEN ];
char m _ LaKey[ KEYLEN ];
char m _ ChnKey[ KEYLEN ];
char m _ ChnMean[ VALUELEN ]; };
```

内存数据用有序表类 OrderList 存储, OrderList 是 CArray 类的继承。OrderList 类记录类型为 KeyUnit 类, 定义如下:

```
class KeyUnit
{
public:
CString m _ Key;
unsigned int m _ Address;
KeyUnit() { };
// Copy constructor required by CArray.
KeyUnit( const KeyUnit& n)
{ m _ Key= n. m _ Key;
m _ Address= n. m _ Address; }
KeyUnit& operator= ( const KeyUnit & m)
{ m _ Key= m. m _ Key;
m _ Address= m. m _ Address;
return * this; }
BOOL AFXAPI operator= = ( const KeyUnit & m) const
{ return ( m _ Key= = m. m _ Key); }
};
```

数据成员: CString m\_Key; unsigned int m\_Address; 分别存储词条和索引, 并且, 对赋值号、逻辑等于等操作符进行重载, 以便 OrderList 类成员函数使用。

2.2.2 算法的运行机制 程序运行开始, 分别将存储三种语言的词条读入内存, 查询时查至相应的词条, 可以获得其索引, 通过这个索引访问存储中英拉中文解释索引记录的文件, 由此可读出相应的中(英、拉)文相关词条即中文解释。

2.2.3 查询算法 “前方一致”算法: 查询首部包含输入词汇的词条, 具体搜索方法采用折半查找算法, 查找效率高。“任意匹配”算法: 查询任何部分包含输入词汇的词条, 具体搜索方法采用遍历算法, 查找效率较低。

## 2.3 接口设计

2.3.1 查询层提供的接口 (1) 得到一个词条的解释数据; (2) 枚举到一个词条的关联词; (3) 在词典中插入自己的新词。

2.3.2 数据库层提供的接口 (1) 得到词条的关联列表; (2) 得到词条的关联词条队列; (3) 得到词条的解释数据入口; (4) 在此条列表中加入新的一项; (5) 在关联列表中加入新的一项; (6) 在关联词队列中加入新的一项; (7) 把相应格式的 Excel 文件转换成数据库文件; (8) 把数据库文件转换成相应格式的 Excel 文件。

2.3.3 用户界面层 Windows 界面显示技术、热键技术、上下文敏感帮助系统等。

2.3.4 数据库层 在运行的时候可以把三张词条表全部载入到内存中去加快访问速度。当然如果内存够大, 还可以把关联列表和关联词条队列文件也载入到内存中, 如果不是全部放到内存中, 可以维护一个 Cache 来提高访问速度。数据库层还包括散列表的散列算法、表的添加、修改、删除, 队列的插入、修改、删除, 相关文件间指针的处理。

## 3 应用条件和环境

树木名称电子词典采用 Visual C++ 编程并编译, 在 Windows (包括 Windows 98, Windows 2000) 环境下运行。计算机硬件最低配置包括 32 兆内存和 5 兆左右硬盘空间。系统安装需要 3.5' 软驱或 5.4' 光驱。

## 4 软件说明

软件界面如图 2 所示, 窗体上部为查询输入部分, 窗体下部为查询结果部分。

窗体标题栏右边为系统功能按钮。从左往右依次为: 第 1 个为帮助按钮, 点击它可以弹出帮助窗口; 第 2 个为菜单按钮, 点击它可以弹出下拉菜单; 第 3 为最小化按钮, 点击它可以退回到桌面右下角的系统任务栏中, 点击其图标查询窗口可以再次弹出。第 4 个为程序关闭按钮, 点击它可以结束程序的运行。

窗体上部编辑框为关键词输入栏, 在此栏输入关键词即可完成查询功能, 此栏具有记忆功能, 能够记忆用户以前输入的查询词汇。在输入栏的后面有两个按键分别为“查询”和“清除”, 点击“查询”按键即可完成查询工作, 点击“清除”按键即可清除输入栏和输出栏内显示的文字。

窗体下部为查询结果和相关设置部分。左边为输入词汇的相关词汇列表, 右边为对应词汇的其他两种语言解释和相关中文说明。相关设置部分可对输入语种和查询方法进行选择。

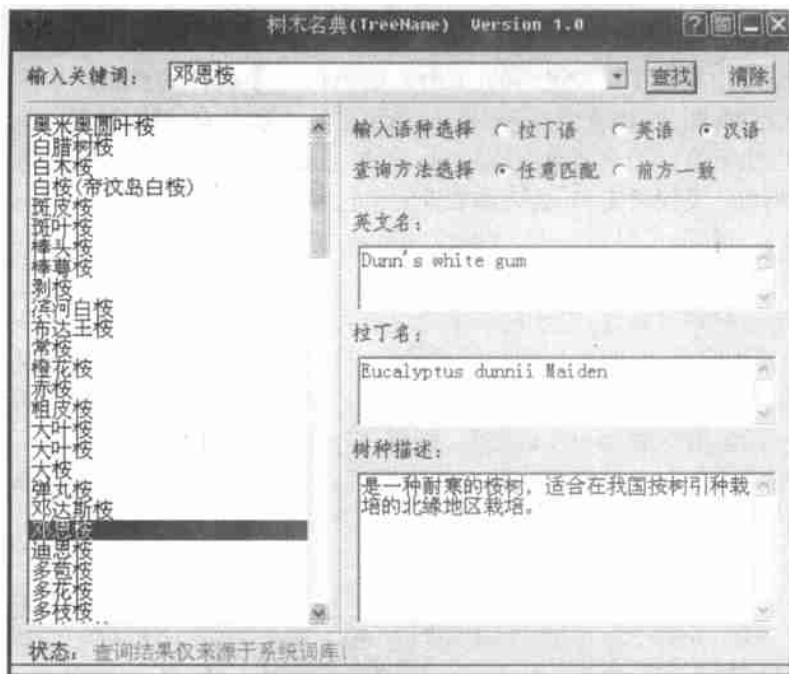


图 2 查询系统界面

## 5 使用方法

将安装盘上的压缩文件解压到指定文件目录, 直接运行 TreeName.exe 即可。也可在桌面建立该程序的快捷运行图标, 每次点击快捷图标即可启动该程序。直接双击应用程序图标, 即可运行树木名典, 软件运行窗口如图 1 所示。点击右上方的最小化按钮, 软件隐藏到电脑桌面右下角的任务栏中, 显示为图标, 再点击图标, 将词典窗体弹出。

树木名典查询用法包括英汉拉、拉汉英、汉拉英对译, 词典窗口界面如图 1 所示。查询工作开始需要对输入关键词的语种和查询方法进行选择, 确定语种和查询方法后, 在词汇输入栏输入所要查询的词条, 然后按下回车(Enter)键, 在窗体左侧部分列表栏中就会出现输入词汇的相关词汇, 并且在窗体右侧出现左侧列表栏中的词汇相关的其他两种语言的解释以及相关的中文说明。窗口上方的输入语种选择项的单选钮表示要查询的词条的语言类型, 你所选择的语种要与输入的关键词语种一致。查询方法项的单选钮表示查询所选择的方法。“前方一致”方法表示查询的相关词汇的首部与输入词汇一致, 而“任意匹配”方法表示查询的相关词汇的任意部位包含输入词汇。

用户通过自定义词库功能建立自己的词库。词汇添加方法: 按词条输入, 即一对一的方式, 输入完毕后, 点击“添加”按钮即可完成更新任务。

使用本帮助系统, 可以使用户快速学会使用本软件。有 3 种方式可以打开本帮助窗体: (1) 点击词典窗体上方标题栏右边的第 1 个按钮; (2) 点击窗体上方标题栏右边的第 2 个按钮, 弹出下拉菜单, 在菜单中选择系统帮助条即可; (3) 快捷键 F<sub>1</sub>。点击帮助窗口左侧帮助主题, 即可在窗口右侧得到相应主题的帮助信息。

## 6 功能扩展与应用前景

本系统在设计中采用了比较灵活的功能模块设计, 利于软件的更新。未来的版本将增加树种的多媒体信息如图片、影音信息等, 此外, 还可更进一步扩充功能, 使其具有树种分类检索表功能。本软件也可直接扩充或改版成植物、动物、昆虫、病原、微生物、菌类等多种生物系统的名称电子词典和分类检索工具, 为诸如农业、林业、生物、医学、环境等相关领域基础研究工作提供非常实用、方便的工具, 是一次突破性的革新。本系统可以很方便地移植到基于网站的互联网平台, 用户就可以实现远程操作, 完成查询工作。

与印刷版的各种词典相比, 电子词典系统具有无法比拟的优点, 它能够及时进行修改、补充, 使系统不断得到完善, 及时根据用户的反馈信息进行改进, 有利于软件质量的提高和功能的完善。也可以在电子版本的基础上, 出版印刷版的词典, 以方便于不能上网或使用计算机的用户。由于电子版具有吸收用户反馈信息、更新、修改和增添词库的显著优点, 在电子版基础上出版的印刷版词典所包括的范围将是以往印刷版本难以比拟的, 它将是一部条目多且最准确的树木名录大词典。

### 参考文献:

- [1] 汉拉英中国木本植物名录编委会. 汉拉英中国木本植物名录[M]. 北京: 中国林业出版社, 2003
- [2] 康定中. 汉英-英汉林业分类词汇和用语[M]. 北京: 外文出版社, 1998. 632
- [3] 中国科学院植物研究所. 新编拉汉英植物名称[M]. 北京: 航空工业出版社, 1996. 1166
- [4] 吴中伦. 国外树种引种概论[M]. 北京: 科学出版社, 1983. 528
- [5] 金山词霸: <http://www.kingsoft.com/>
- [6] 网际金典: <http://www.roboword.com.cn/>
- [7] 东方快车: <http://dlkc.sunw.com/index.html>

## Development of an Electronic Dictionary of English, Latin and Chinese Names of Tree Species

*ZHENG Yong-qi*<sup>1</sup>, *ZHANG Chuan-hong*<sup>1</sup>, *ZHENG Hong-tao*<sup>2</sup>, *ZHENG Zhi-hua*<sup>2</sup>, *LI Bo-ying*<sup>1</sup>

(1. Research Institute of Forestry, CAF, Beijing 100091;

2 Beijing University of Aeronautics and Astronautics, Beijing 100082)

**Abstract:** An electronic dictionary of English, Latin and Chinese names of tree species (TreeName 1.0) was developed, providing searches for names of tree species, genus and family in any of the three languages. The electronic dictionary contains thousands of names of tree species and brief species descriptions, providing a rapid, accurate and efficient powerful tool as a dictionary for those working with forestry and plant sciences. The software was developed in Visual C++ , using dialogue boxes as the search interface and file system based databases as the framework of the software. Flexible function modules were used in the design. In addition to its strong capacity of providing searches, the software was also designed to allow users to create or update their own customizable database of names of tree species. The software can be extended to cover names of plants, animals, birds, insects and any other biological organisms. The software provides incomparable advantages over the traditional printed dictionary.

**Key words:** tree species names; electronic dictionary; search; database