

中国畜禽品种资源网络数据库系统的构建

熊本海, 马月辉, 罗清尧, 庞之洪

(中国农业科学院畜牧研究所, 北京 100094)

摘要: 在规范化《中国畜禽品种志》和重新调查搜集到的大量畜禽种质资源信息后, 利用 TBF 全文信息非结构数据库, 构建了“中国畜禽品种资源网络数据库”, 并通过 Internet 以 <http://www.cdad-is.org.cn/> 网址发布。在网站内系统提供了包括全文检索、组合检索、二次检索、字段检索和多库检索等功能, 并指出了本网络数据库的今后发展目标是如何获取我国畜禽品种资源的动态信息, 以及实现资源信息从内容管理到知识管理的逐步过渡。

关键词: 畜禽; 资源; 网络数据库

中图分类号: S813.9

文献标识码: A

文章编号: 0366-6964(2005)01-0043-05

据世界粮农组织近 10 多年来对 170 个国家记录在案的 6 379 个家养畜禽品种的调查结果, 在近 100 年里已有 740 多个品种绝灭, 1 335 个品种处于高度濒危状态。据专家预测, 如不采取措施, 还将失去 2 000 个畜禽品种^[1]。我国在“六五”-“七五”期间, 在 20 世纪 70-80 年代以郑丕留教授为首的中国畜禽品种调研组组织全国同行历时 7 年对全国畜禽品种进行了全面系统的调查, 并已出版了《中国猪品种志》、《中国牛品种志》、《中国羊品种志》、《中国马驴品种志》、《中国家禽品种志》。根据该品种志载, 截至 1986 年有 194 个地方品种、45 个培育品种、43 个引进品种^[2-6]。从“八五”至今, 中国农业科学院畜牧研究所“畜禽遗传资源研究室”一直承担和坚持我国畜禽种质资源动态信息调查工作, 共调查了全国 28 个省区的 459 个畜禽种质资源信息, 基本掌握了各个品种的濒危程度和群体数量变化趋势。一个值得注意的问题是这批品种资源正在“洋种”即“良种”的指导思想上, 不断被杂交“改良”淘汰。目前国内规模化养殖业中所用畜禽品种基本上是从国外引进品种或品系及其杂交后代。“有钱引种”, 无钱育种、保种的现象仍然普遍存在。

依据我国畜禽种质资源状况的动态变化特征, 及早建立中国家养动物品种资源数据库, 可以盘活、再现、分析和充分利用已经收集到的资源信息, 及时为国家或区域性制定保种、引种方案提供科学依据, 对提高政府部门决策能力、业内人士对资源状况的

全面掌握和社会公众对资源的认识, 将会起到积极的作用。特别是, 如果能够通过现代信息技术, 建立完善的数据采集、远程维护更新体系, 则能及时对我国畜禽资源品种的濒危程度作出由量变到质变的预警分析, 以避免“亡羊补牢, 为时晚也”的尴尬局面。

1 中国畜禽品种资源网络数据库系统平台构建原理与方法

1.1 数据网络操作系统

服务器的操作系统选定为 Windows 2000 Advanced Server, 即 Windows NT 的升级版本, 含有信息服务器 IIS 5.0 系统, 负责 Web 站点的管理与信息发布。客户端可为 Windows 2000 Professional, Windows 95, Windows 98, Windows ME, Windows XP 中的任意一种。

1.2 数据库管理系统

由于畜禽品种资源信息特点是以描述性文本为主, 并含有图片、影像等多媒体信息, 也包括部分数值型数据, 因此本系统选用了 TBF 全文信息库^[7]。该类数据库通过采用多种压缩算法, 将结构化数据库与非结构化全文信息库完美地结合在一起, 数据可变长储存, 最大限度节省空间, 为实现包含图片及多媒体信息的数据库毫秒级查询提供技术支撑。本系统的所有数据资源全都存放在如图 1 所示“全文信息服务器”中, 可通过“库管理(TBS Client)”和“文件管理(FTS Client)”终端进行远程数据库和文件维护。网络用户通过 ODBC 接口, 实现对不同格式数据的统一处理与访问。

收稿日期: 2003-07-10

基金项目: 科技部科技基础性项目(2001DEA10006)

作者简介: 熊本海(1963-), 男, 湖北红安人, 博士, 副研究员, 主要从事畜牧信息技术研究。

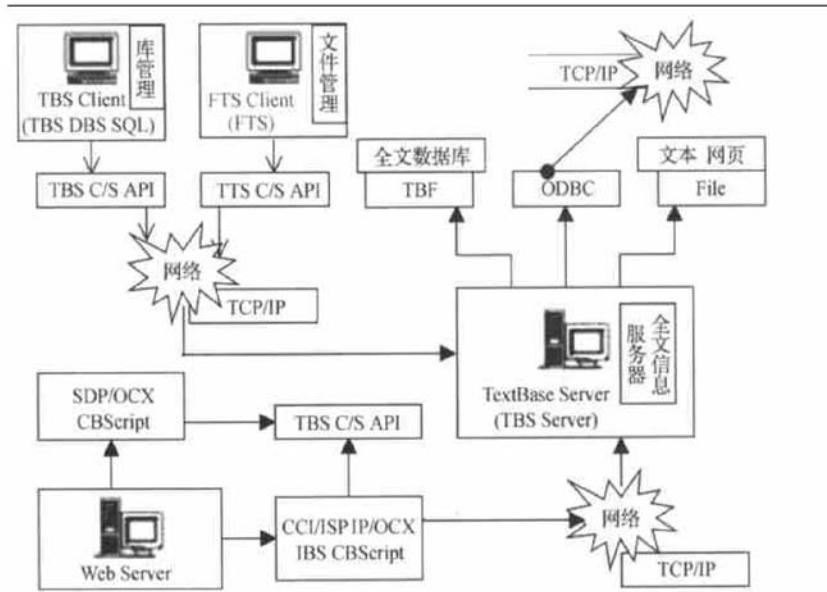


图 1 中国畜禽品种资源网络数据库系统软、硬件结构

Fig. 1 Software and hardware frame of network database on Chinese Animal Breed Resources

1.3 系统结构与动态网页技术

如图 1 所示, 客户浏览器与 Web 服务器之间为 Browser/Server 结构, Web 服务器与全文数据库服务器之间为 Client/Server 结构, 这种三层结构 (C/S) 可实现分布式动态负载均衡、多机并行检索、超大容量、多语种、多媒体、高可靠性的信息全文检索系统。页面编程语言采用 CScript 脚本语言。它的使用方法与语法风格类似于 ASP 中的 JavaScript。

1.4 数据源

1.4.1 国家确认的品种资源信息 主要依据《中国猪品种志》、《中国牛品种志》、《中国羊品种志》、《中国马驴品种志》、《中国家禽品种志》。根据该品种志载, 截至 1986 年我国有 282 个畜禽品种。家畜品种 (含类群) 共 207 个, 其中猪 66 个, 山羊 23 个, 绵羊 30 个, 马 33 个, 牛 37 个, 牦牛 5 个, 水牛 1 个, 驴 10 个。而家禽品种 (含类群) 共 75 个。其中鸡 47 个, 火鸡 1 个, 鸭 14 个, 鹅 13 个。

除国家品种志提供的数据外, 还有相当一部分品种资源信息来自各省市公开发表的地方畜禽品种志^[8-20]。按品种分有猪 18 个, 山羊 12 个, 绵羊 3 个, 马 4 个, 牛 10 个, 牦牛 1 个, 水牛 8 个, 大额牛 1 个, 驴 1 个, 骆驼 6 个, 兔 4 个, 总共 68 类或类群的畜种信息; 同时还收集了鸡品种 24 个, 火鸡 1 个, 鸭 12 个, 总共 37 类或类群的地方禽品种资源信息。

对已发表的资源信息的处理重在内容的标准化处理, 便于信息的标识、入库和索引。

1.4.2 尚未确认的品种资源信息 结合中国农业科学院畜牧研究所自 1999 年承担的科技部科技基础性项目“畜禽种质资源收集、整理、保存、评价”以来的动态调查信息, 进一步扩充和更新了我国畜禽种质资源数据库。共调查了全国 28 个省区的 459 个畜禽种质资源信息, 其中猪 105 个, 牛 78 个, 马驴 53 个, 绵山羊 104 个, 家禽 119 个; 并发现已经灭绝的种质资源有 9 个。其数量变化趋势为: 177 个品种的群体数量上升, 219 个品种数量下降, 3 个品种数量不变, 51 个品种没有以前的群体数量记录而无可比性。受严重威胁的品种有 44 个, 受较大威胁的品种 36 个, 受威胁品种 14 个, 受潜在威胁品种 58 个, 相对安全的品种 298 个^[21,22]。这些信息是对《中国畜禽品种志》最新、最全面的补充。

1.5 信息检索系统的概念

信息检索是畜禽遗传资源网络数据库构建的重要内容。本系统在“资源数据库”基础上, 通过全文信息服务器整合各种数据资源。本系统涉及的主要检索方法的定义及其特性如下:

全文检索: 用户随时输入一个或若干个字、词, 就可从全文库中检索出包含此字、词的全部记录。通过下拉按钮, 用户可以选择显示的浏览字段和控制显示检索结果在每屏出现的记录条数。对检索结果还可进行二次检索, 以缩小检索范围, 提高检索效率^[7]。

字段检索: 用户可针对某个字段检索, 凡是字段建立“字段检索”的, 都可以进行字段检索。在输入

检索词的同时,可限定检索字段(此字段需建立字段索引)。检索方式可以是包含和相等的关系,对检索出的结果同样可进行二次检索。

组合检索:组合检索是全文检索和字段检索的结合,是在限定多个字段的基础上实行的全文检索。可通过指定字段条件来对某个字段或多个字段的条件下进行全文检索。

多库检索:多库检索是一种综合检索方式,可以进行跨库检索。用户可以选择几个或全部数据库,从中检索一个或几个字、词、句。对检索出的记录按库名分开显示。如果一个全文库中的字段之间有一种树形层次关系,可建立多级索引,通过多级索引可逐级缩小检索范围,实现快速检索。

依据以上对检索系统中各功能模块内涵的界定及其具体程序的实现,从而构成《中国畜禽品种资源网络数据库》的检索查询系统,为在网络前端快速有效地响应客户的各种需求提供了可能。

2 结 果

2.1 系统网站的建立与检索结果分析

按照第 1 部分采用的网络平台、数据库系统、页面语言和数据源,我们构建了“畜禽品种资源网络数据库”,并通过 Internet 以 <http://www.cdac-is.org.cn/> 网址发布。

进入主页后,系统直接提供不同的查询或浏览入口,以满足不同的检索要求。

2.1.1 按畜禽类别进入查询或检索 在主页左边从“中国家畜品种志”栏目下,选定“牛(cattle)”进入后,首先可以看到来自《中国畜禽品种志》中所有品种的牛共 67 个,每页显示 10 个,共分 7 页显示完(图 2)。若查看某个品种的具体资源数据,只要点击显示出的图片即可。在左边显示可进一步进行检索的字段名称,或从该窗口进入“全文检索”或“组合检索”。



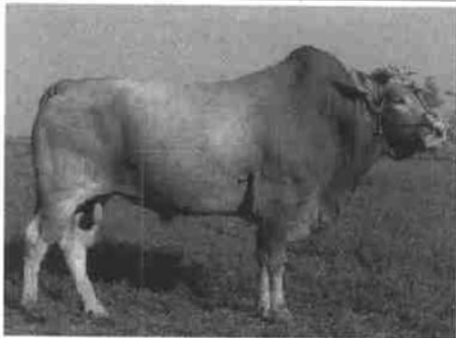
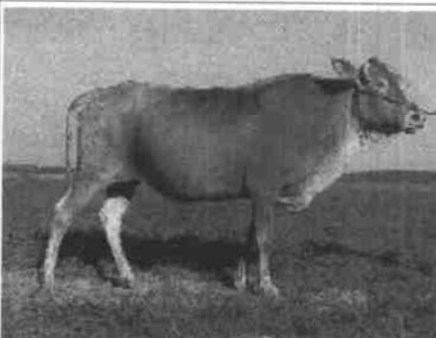
国家动物品种志		请输入检索词: _____ <input type="button" value="确定"/> <input type="button" value="重输"/> <input type="button" value="关闭窗口"/>	
<ul style="list-style-type: none"> ∴ 品种登记号 ∴ 畜种名称 ∴ 品种来源 ∴ 品种名称 ∴ 品种英文名 ∴ 品种别名 ∴ 类群名称 ∴ 类群别名 ∴ 浏览信息 ∴ 全文检索 ∴ 组合检索 		检索词: 共检中: 67 条, 当前页: 1/7 2 <input type="button" value="跳转"/>	
	序 号	品种 名称	类群 名称
	1	秦川牛	
	2	南阳牛	
		照片(公)	照片(公)
			
			

图 2 来自国家家畜品种志牛品种查询的初步结果

Fig. 2 The first searching results from the Cattle Breed Records of China

如果按“畜种名称”进入查询,则实质上是按畜种名称进行归类查询,查询结果为:大额牛(1)、牛(47)、水牛(12)和牦牛(7)。括号中表示品种数。若按“品种来源”查询,其结果为:地方(56)、培育(4)和引进(7)。由此可见,如果一个全文库中的字段之间有一种树形层次关系,在建立多级索引后,通过多级索引可逐级缩小检索范围,实现快速检索。

2.1.2 按数据库进入查询或检索

图3 中国家禽品种志数据库所含数据表

Fig. 3 The included data tables in the Poultry Breed Records of China

2.1.3 按“中心产区”进入查询或检索 如主页中间所示,系统提供了2种途径进入按“中心产区”的查询或检索:一是按主页中间中国地图指示的省份所处位置,点击后,系统直接将畜禽品种“中心产区”在本省市的所有品种(或类群)搜索出来。所以检索出的品种不仅包括国家或地方品种志的,而且也包括不同的家畜和家禽品种,然后再进行全文检索、组合检索、字段检索和二次检索等,了解宏观与微观的资源信息;二是按主页右上角指示的省市名称进入查询。2种方式的检索结果是完全一样的。

3 讨论

通过现代信息技术,将过去和现在收集到的有关我国畜禽品种资源的数据经过格式化处理后,形成了网络化数据库,实现了资源信息的内容管理,并借助Web技术,可以为国家政府决策部门、业内人士和社会公众提供我国畜禽资源信息的查询服务,这是本系统构建最基本的目标,但不是最终目标。本系统建设的最终目标应主要体现在以下方面:

实现我国畜禽种质资源的动态信息管理,及时反映我国畜禽种质资源的现状,如品种的群体数量的变化趋势、濒危程度和资源开发利用状况等,为国

如主页中间所示,系统提供了按指定数据库(下拉弹出菜单)进入查询。这里指示的数据库可包含数量不等的数据库组成,而对输入的关键词的检索,则是在指定的数据库所包含的数据表中进行的与关键词的匹配检索。例如,图3显示了“中国家禽品种志”数据库所含数据表。限于篇幅,对检索效果不作具体说明。

家和地方制定畜禽种质资源保护措施提供科学依据。因此,必须建立切实可行的畜禽种质资源信息采集和数据更新、维护机制,保证变化后的资源信息能及时反映到系统中来。这不仅仅是系统数据维护的技术问题,更重要的是资源数据的归属、上传、享用和使用权限等数据政策问题,需要涉及科研、政府和地方有关资源管理部门的相互协调,可通过建立数据中心(国家)和数据站(各省市),明确分工,各享义务与责任,改变传统的、静态资源信息的收集模式为网络在线信息更新手段,提高我国资源信息的时效性。

实现我国畜禽品种资源信息的知识管理,及时对我国畜禽资源品种的濒危程度作出由量变到质变的预警分析。就目前而言,即使系统提供了动态的种质资源信息和丰富的信息检索模式,只能是对信息的内容管理,目的是为了“内容”能够在正确的时间,以正确的形式传递到正确的地点和人。而知识管理则是对相关信息资源进行过滤、提炼及加工而得到的有用、有价值资料。知识管理的任务是对资源信息的显性知识和隐性知识进行处理,并把这些知识用一种适合于用户理解的方式表示出来,提供给业内共享、吸收、利用。实现资源信息的知识管

理不仅需要大量畜禽品种的资源数值信息,而且也需要大量的动物资源信息学基础知识。在此基础上,结合“数据库中知识发现(Knowledge Discovery in Databases, KDD)”技术^[23],才能真正使已有的畜禽品种资源数据信息增值,如进行各种畜禽品种资源存量的预警分析,应成为本系统重要的功能目标。

参考文献:

- [1] 张子仪. 我国养殖业在新世纪中的若干难题与对策刍议[J]. 国外畜牧科技, 2001(1): 3~ 5.
- [2] 中国家畜家禽品种志编委会. 中国猪品种志[M]. 上海科学技术出版社, 1986~ 1989.
- [3] 中国家畜家禽品种志编委会. 中国牛品种志[M]. 上海科学技术出版社, 1986~ 1989.
- [4] 中国家畜家禽品种志编委会. 中国羊品种志[M]. 上海科学技术出版社, 1986~ 1989.
- [5] 中国家畜家禽品种志编委会. 中国马驴品种志[M]. 上海科学技术出版社, 1986~ 1989.
- [6] 中国家畜家禽品种志编委会. 中国家禽品种志[M]. 上海科学技术出版社, 1986~ 1989.
- [7] 熊本海, 庞之洪, 罗清尧, 等. 全文检索技术在国家饲料数据中心网站的应用[J]. 中国饲料, 2002, 23: 25~ 26.
- [8] 贺晓兰. 湖南省家禽家畜品种志和品种图谱[M]. 湖南科学技术出版社, 1984.
- [9] 涂友仁. 内蒙古家禽家畜品种[M]. 内蒙古人民出版社, 1985.
- [10] 甘肃省畜牧厅. 甘肃省家禽品种志[M]. 甘肃人民出版社, 1986.
- [11] 广西家畜家禽品种志编委会. 广西家畜家禽品种志[M]. 广西人民出版社, 1987.
- [12] 陈永泽. 贵州省家禽品种志[M]. 贵州科技出版社, 1993.
- [13] 吕效吾. 山西省家畜家禽品种志[M]. 华东师范大学出版社, 1984.
- [14] 李永绿. 广东省家畜家禽品种志[M]. 广东科学技术出版社, 1987.
- [15] 云南省家畜家禽品种志编委会. 云南省家畜家禽品种志[M]. 云南科技出版社, 1987.
- [16] 福建省家畜家禽品种志编写组. 福建省家禽家畜品种志和图谱[M]. 福建科学技术出版社, 1985.
- [17] 李哲民. 黑龙江省家畜家禽品种志[M]. 黑龙江人民出版社, 1985.
- [18] 湖北省家畜家禽品种志编委会. 湖北省家畜家禽品种志[M]. 湖北科技出版社, 1985.
- [19] 邱祥聘. 四川省家畜家禽品种志[M]. 四川科学技术出版社, 1987.
- [20] 陕西省家畜家禽品种志编辑委员会. 陕西省家畜家禽品种志[M]. 三秦出版社, 1988.
- [21] 马月辉, 徐桂芳, 王端云, 等. 中国畜禽遗传资源信息动态研究[J]. 中国农业科学, 2002, 35(5), 552~ 555.
- [22] 马月辉. “畜禽种质资源收集、评价、保存”工作简介[J]. 中国农业科技导报, 2003, 5(3): 80.
- [23] 夏敏华. 内容管理创造数字财富[N]. 中国计算机报, 2002(1124): 2~ 3.

Construction of Network Database on Animal Genetic Resources In China

XIONG Ben-hai, MA Yue-hui, LUO Qing-yao, PANG Zh-hong

(Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 10094, China)

Abstract: This study has set up a network database on animal genetic resources in China depending on standardized resources data from bath “Chinese Animal Breed Records” and much collected animal genetic information anew, software and hardware circumstance and TBF non-structural full-text database. At the same time, we have issued our website <http://www.cdad-is.org.cn/> by Internet. Inside the website, we supply many searching models such as full-text searching, field searching, combination searching, twice searching and multi-databases searching etc. Finally, it points out the future development objectives of the network database as follows: The first is how to get dynamic animal genetic information of China, the second is to realize transition of information serving from current content management to future knowledge management.

Key words: animal; resource; network database