

中南半岛药用植物资源信息系统建设*

□ 刘海波 彭 勇** 张新国 肖培根

(北京协和医学院药用植物研究所 北京 100193)

龙 立 刘延淮**

(北京东方灵盾科技有限公司 北京 100191)

摘要:“中南半岛药用植物资源信息系统建设”是2009年科技部对发展中国家的援助项目,旨在对中南半岛地区药用植物资源进行系统地信息收集,一方面对该地区的国家进行技术和经济的支援,另一方面,为我国对该地区药用植物资源的开发奠定信息基础。本文详细介绍了该项目的承担单位、项目目标与内容,并对中南半岛药用植物数据库的设计、辅助录入系统的开发、数据收集、机器翻译、信息平台的建设方面取得的成果进行论述,希望对国内药用植物资源研究人员起到一定借鉴作用。

关键词:药用植物资源 数据库 中南半岛

doi: 10.3969/j.issn.1674-3849.2010.06.015

药用植物资源是传统医药可持续发展的基础,是我国发展优势生物医药经济、保障人民健康的重要战略资源。但长期过度开发、生态环境恶化等因素导致全世界20%以上药用植物资源濒临灭绝。我国情况尤为严峻,初步统计表明我国3000种濒危植物中药用植物占60%。大量中药种质资源濒于灭绝,如野生人参、野生当归等^[1]。特别是近年随着中药产业的迅猛发展,每年都有数10种中药资源被列入濒危物种名单,已成为国内外瞩目的热点问题^[2-3]。

中药种质资源濒危和流失,已严重制约了中医药可持续发展。保护好现有国内中药种质资源,并积极探索其他国家的药用植物资源,促进药用植物引种和

药材贸易,是我国中药产业可持续发展的必由之路。

东盟10国是世界上重要的经济合作区域之一,与中国有着重要的经济贸易往来。2008年中国医药出口成交总额为31.73亿美元,其中对东盟出口24.07亿美元,占总额的75.7%;进口成交总额是15.17亿美元,其中从东盟进口6.59亿美元,占总额的43.4%。东盟已经成为中国医药出口第三大贸易伙伴,仅次于美国和日本^[4]。

东盟中越南、老挝、缅甸、泰国、柬埔寨5国同处于中南半岛,在地理上与中国接壤,植物多样性非常丰富,而且植物类型与中国西南部省份比较接近,比较利于驯化引种;在历史上,中南半岛5国与中国有着悠久的交流历史,原产于这些国家的血竭、沉香、檀香等,现在都已经成为中药里的常用药;在政治

收稿日期:2010-03-18

修回日期:2010-04-28

* 科技部国际合作司对发展中国家科技援助项目:中南半岛药用植物资源信息系统建设,负责人:刘延淮。

** 通讯作者:彭勇,研究员,主要研究方向:中药资源学与中药信息学研究,Tel:010-62894462,E-mail:ypeng@implad.ac.cn;刘延淮,研究员,原国家知识产权局知识产权出版社专利数据研究中心主任,主要研究方向:专利信息,Tel:010-51296683,E-mail:tyh@eastlinden.com。

[World Science and Technology/Modernization of Traditional Chinese Medicine and Materia Medica] 924

上,近 10 年来,这 5 个国家政府与中国都保持着比较好的睦邻友好合作关系,比较利于进行科研交流;在经济和科学方面,这 5 个国家都属于发展中国家,经济和科研水平相对落后,中国对这 5 国的技术援助能够对这些国家的经济、科学发展产生比较大的推动作用,影响力比较大。综上所述,中南半岛 5 国是探索中国以外天然药物资源的首选区域。

2009 年北京东方灵盾科技有限公司承担了科技部对外合作司对发展中国家援助项目《中南半岛药用植物资源信息系统的建设》,合作单位包括中国医学科学院药用植物研究所和越南卫生部药材研究院。该项目旨在广泛地收集东南亚区域内的天然药物资源及传统使用方法等信息,为我国与这些国家合作进行药用资源的收集、引种、开发等工作提供信息储备,并对这些国家的传统药物研究产生积极的带动作用,有着重要的科学、经济和政治意义。

本文对该项目的实施过程和取得的成果进行简要介绍。

一、承担单位

项目主要承担单位北京东方灵盾科技有限公司在信息数据库建设尤其是在专利信息数据库建设方面是国内顶极的企业,在国家科技部和国家知识产权局项目支持下,先后建设了多个深度加工标引的专利信息数据库,同时积极参与科技部国际合作司的“中医药国际科技合作项目”,在天然药物和世界传统药物专利信息深度加工标引技术和检索系统开发方面,有多年的研究经验,并与诸多国内外专家保持着广泛的合作关系。

中国医学科学院药用植物研究所是国内顶尖的从事药用植物资源研究专业机构,承担着中草药物种资源保存、珍稀濒危物种保护、中药材规范化种植(GAP)的制定与实施、中药材标准对照品的研究、中药资源普查等重大科研任务,长期以来与发展中国家有着广泛的合作关系。在本项目中,作为与国外天然药物研究机构的桥梁,起到重要的联系和沟通作用。

本项目的合作国家越南,位于中南半岛的东侧,国土狭长,有漫长的海岸线,是中南半岛 5 国中纬度跨度最大的国家,国土的 3/4 是高原和山地,平原只占 1/4,因此植物物种资源非常丰富。在历史上由于长期受中国的影响,有悠久的传统药物使用

历史,积累了大量药用植物利用的知识^[5]。但是,由于近现代长期处于法国殖民统治和战争状态,科研水平发展较低,特别是信息技术比较落后,因此对本国丰富的药用植物资源缺乏数据采集和利用能力。本项目的重要目的之一,就是帮助越南对本国药用植物资源数据进行系统化和信息化,建立现代科学研究所要求数据库和信息系统,推动该国传统药物研究的发展,促进中越科研和经济的交流,巩固两国的友好关系。

国外合作单位越南卫生部药材研究院是越南卫生部直属机构,历史悠久,在越南天然药物和传统药物的研究方面有较强的研究实力,积累了一定的文献和数据资源,是本项目适合的国外合作伙伴。

二、目标与内容

“中南半岛药用植物资源信息系统”项目目标是以高质量的中国中药信息资源整理加工标准、规则和计算机处理技术为基础,建设数据质量最高、检索功能全面的药用植物信息系统,拓展科研人员信息交流空间,促进中国和中南半岛国家药用植物研究或天然药物研究事业的发展。

以获取中南半岛药用植物资源的第一手资料,结合现有中国药用植物数据和网络技术资源,搭建中南半岛药用植物资源信息系统。项目内容主要包括:收集中南半岛药用植物资源方面的信息,内容涵盖植物学、药理学、植物化学、生产贸易等方面,主体内容实现中、英文双语种;建立中南半岛药用植物资源信息检索平台;在北京和河内建立信息中心,对外提供在线检索服务;与越南等中南半岛国家在药用植物研究领域建立的长期合作关系。

三、完成情况

1. 中南半岛药用植物数据库的设计

传统药物研究是一个多学科交叉的领域。本项目涉及植物学、化学、药理学、专利信息等一系列数据类型,每种数据有自身的特点,又同时与其他数据联系。只有在前期设计出一个具有开放性、健壮性的数据库结构,才能保证之后的数据收集工作有据可依。图 1 是我们设计的中南半岛药用植物数据库数据结构。该数据库的数据主体包括植物学和化学数据、传统药物数据、传统药物相关专利数据 3 个部分。3 部分信息靠植物学分类学编号和药物登记号联

接到一起。该设计较好的解决了不同类型数据的独立性和关联性的矛盾，最大程度上降低了数据的冗余度，提高了存储和检索效率。

分类学登记表的记录单位为分类学上的一个种(细化到变种、亚种)，存储植物分类学和植物学信息。天然产物登记表用于记录药用植物活性成分的各种属性。传统药物信息部分是主要需要收集的数据。由于中国和中南半岛国家的信息结构不同，因此存储于不同的数据表，有针对性地设计了不同的字段。每条记录对应一种传统药物，根据天然药物的不同药用部位、不同炮制方法处理和不同使用方法，分别单独进行记录。中国以外的传统药物基本都配有当地语言的名称(使用本地语言或者用发音标注)，基于此实现数据库的多语种查询功能。天然药物专利数据库是东方灵盾公司自主研发的传统药物专利数据库，存储了大量经过人工深加工的传统药物专利文献及其相关信息，该数据库通过转库检索的形式与中南半岛药用植物数据库链接。

2. 辅助录入系统的开发

中南半岛药用植物数据库的建设是一个大数据量的工程，如果完全靠人力进行数据录入工作，工作量大且容易出错。因此，在整个工作开始前，专门制订了完善的工作流程，并开发了专门的辅助录入软件系统，完成整个工作中大部分的重复性、机械性工作。整个流程由3部分构成：①人工标引规则及半自动标引软件(图2)；②自动入库系统；③后期自动勘误系统。事实证明，该套系统在整个数据工作中发挥了巨大的作用，极大的提高了工作速度、降低了错误率，达到了降低了项目人工成本、节省时间的目标。

3. 数据收集

对中南半岛各国药用植物信息的收集是整个项目的工作重点，也是最富成果的一个方面。中国医学科学院药用植物研究所和越南药用植物研究院两家合作单位分别给予了大力支持，前者提供了大量外文文献资料，供数据采集使用，并在数据整理方法上进行了科学指导；后者直接派人参加了数据加工工作，根据项目的数据要求，整理加工了161种越南药用植物的资料，另外，还有248张照片，包括了植物的天然形态、种植

方法、成药形态等信息。

本项目的数据来源主要有相关的图书文献、越南方面提供的数据和东方灵盾原有的数据积累。3类数据各有特点，因此需要不同的后期处理方法。越南方面提供的数据由于先期制订了数据采集标准，因此数据规范性比较好，经过格式转换清洗后就直接入库。东方灵盾原有数据资源比较丰富，但与本项目的侧重点不同，因此需要进行规范化整理。图书文献的数据量非常大，处理也非常繁琐，我们特别开发了专门的自动化标引软件进行图书文献的收集工作。3个来源的数据经过清洗、融合和人工校对后，形成最终的数据库。

经过一年的不懈努力，本项目取得了丰硕的成果，最终收集的数据量达到并超过了项目任务书中

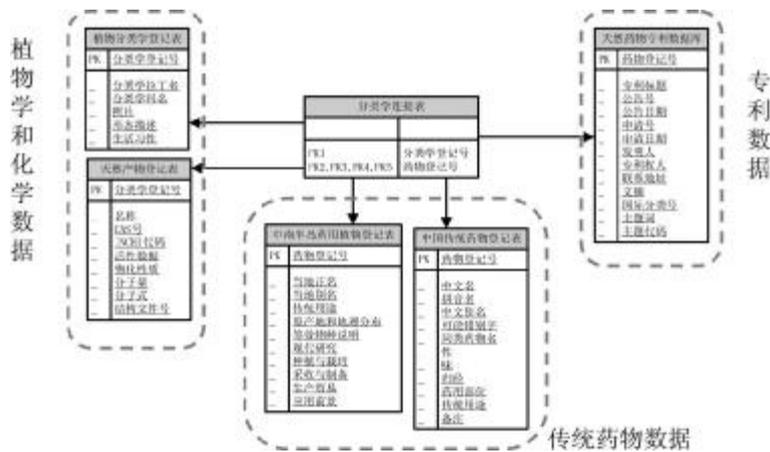


图1 中南半岛药用植物数据库的数据架构

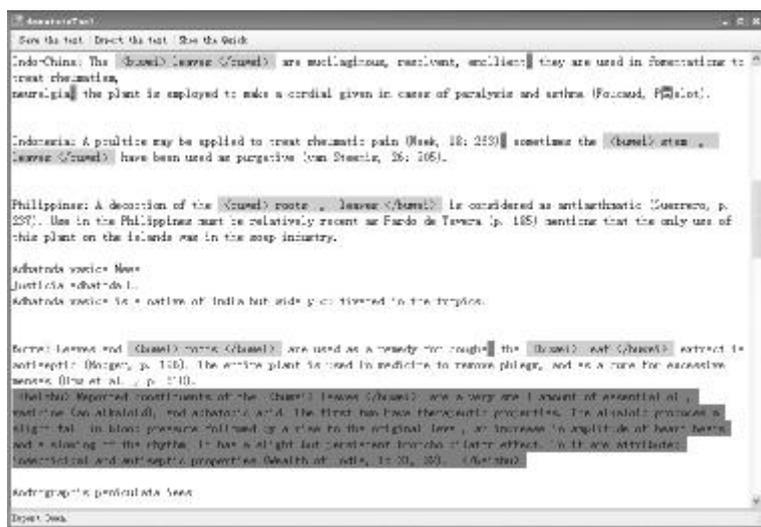


图2 专门开发的半自动标引软件

设定的目标。共收集药用植物 12977 种,天然产物超过 5000 种,天然药物相关专利超过 11 万,所有内容实现中英双语种。

所收集的药用植物数据共 453 科、2831 属、12977 种,各地区传统用法 14917 种,植物不同部位的药用方法 20230 种。专利方面收集了 1985 年以来世界上 22 个国家 2 个国际组织以中药为核心的所有天然药物及其提取物方面的专利信息。基本上形成了一套目前国内系统性最高、质量最好的中南半岛地区药用植物资源资料库,达到了预期的目标。

在地域方面,所搜集的植物中在中国有分布的最多,达到 9687 种,分布在中南半岛,以及马来半岛、印尼、台湾、菲律宾等地区的植物共 4499 种,其中有 1209 种在中国和东南亚地区都有分布。在内容方面,4571 种药用植物有英文的分布、形态、药效、种植等方面的信息。1350 种药用植物附有不同国家地区的名称,其中 161 种由越南药用植物研究院提供了越南文的名称。

在规定的数据类型之外,我们还针对比较重要的药用植物,收集了图片 1049 张,包括野生形态、栽培状态、药用部位等方面的照片,以及手绘的植物形态图片等等,方便了未来研究者的使用。

4. 机器翻译

采集数据的机器翻译是项目后期的主要工作之一。所收集的数据多是单语种的,通过机器翻译和人工校对实现了数据库的中英双语种。翻译工作基于东方灵盾自有的机器翻译系统,该系统采用当前最先进的基于统计的机器翻译技术、结合模板和记忆技术、并将专业领域细分,同时采用海量的专业术语支持确保翻译的质量,并不断地积累语料,重新训练系统,提高翻译系统性能和翻译结果的准确率。

5. 信息平台的建设

在数据平台的架构设计方面,为了保证最高的检索效率,我们将数据分为两部分存储,传统药物和植物学部分信息存储于客户端,专利和化合物数据存储于网络服务器,这样既加快了用户一般使用时的检索速度,又降低了网络数据库的访问压力。同时,我们在后台进行持续的数据加工工作,并定期通过

网络数据库对用户数据进行更新,保证用户及时得到最新的数据(图 3)。

本项目信息系统的界面设计以友好、便捷为原则,采用目前最流行的 Google、百度等通用检索引擎的模式,用户在不需要培训的情况下直接可以使用,在最短的时间内检索得到最需要的信息,检索结果图文并茂、一目了然(图 4)。目前,在北京的数据系统已经可以投入使用*,与越南方面合作的河内信息中心也在建设当中。

四、结 语

本项目以毗邻中国的中南半岛 5 国为突破口,以国际合作为纽带,一方面收集中南半岛地区药用植物资源,为我国对外药材贸易和国外药用植物资源引种等工作打造信息基础。另一方面,带动中南半岛发展中国家的传统药物研究进展,增加中国在该地区的影响力。

在科学和经济方面,本项目产生的效益包括:

第一、对中南半岛的药用植物资源进行了比较系统的信息收集工作,形成了国内第一套较完整的中南半岛药用植物数据集,对今后开发中南半岛地区药用植物资源有很高的参考价值。

第二、拓展信息交流的渠道,发现商机,增加传统药材贸易的种类和规模。

第三、参考周边国家对药用植物利用的方法,深入挖掘与中南半岛临近省份的药用植物资源,发现

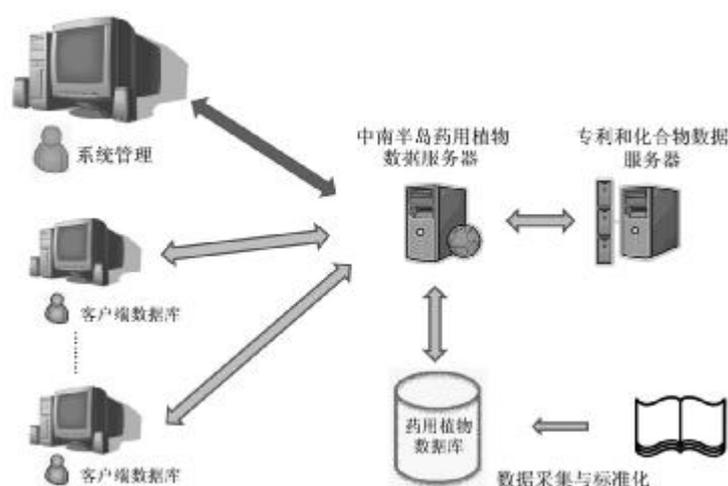


图 3 中南半岛药用植物数据库的系统架构图

* <http://www.wtmpd.com/>



图 4 中南半岛药用植物数据库的检索界面及结果展示

新用途,开发新产品。

第四、能够对东亚植物区系中的药用植物的药效特点、分布等作总体上的研究,并与欧洲、非洲、美洲的药用植物资源作宏观上的比较,挖掘知识、发现规律。

参考文献

- 1 姚振生. 珍稀濒危药用植物保护的探讨. 江西林业科技, 1995(5):38-39.
- 2 肖小河, 肖培根, 王永炎. 中药科学研究的几个关键问题. 中国中药杂志, 2009, 34(2):119-123.
- 3 肖培根, 肖小河. 21 世纪与中药现代化. 中国中药杂志, 2000, 25(2):67-70.
- 4 周杰, 刘玫君, 季益宇. 东盟中医药——潜力巨大的市场. 国际商报, 2009年6月30日.
- 5 张新国, 彭勇, 阮友蓉. 越南药用植物研究与开发概况. 亚太传统医药, 2009, 5(9):6-9.

Indo-China Peninsula Official Plant Resource Database

Liu Haibo¹, Peng Yong¹, Zhang Xinguo¹, Xiao Peigen¹, Long Li², Liu Yanhuai²

(1. Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Beijing 100193, China;
2. Beijing East Linden Science & Technology Co., Ltd., Beijing 100191, China)

Abstract: Sponsored by the Ministry of Science and Technology of China, "INDO-CHINA PENINSULA MEDICAL PLANTS DATABASE BUILDING" is a technology-supporting program for the developing countries around China in 2009, aiming for a systematic data collection of the officinal plant resource in the Indo-China peninsula. The data is expected to facilitate the cooperation in the traditional medicine between China and the countries in the Indo-China peninsula. This paper presents a brief introduction regarding the program progress, which would serve a reference for other researchers.

Keywords: Officinal plant resource, Database, Indo-China peninsula

(责任编辑:张志华 李沙沙, 责任译审:张立巍)