

# 《海洋生物分类检索系统》功能模块的研究

杨俊杰<sup>1</sup>, 胡建廷<sup>1</sup>, 刘保忠<sup>2</sup>, 盖明礼<sup>1</sup>, 苏文<sup>1</sup>, 赵先庭<sup>1</sup>

(1. 山东海洋工程研究院, 山东 青岛 266071; 2. 中国科学院 海洋研究所, 山东 青岛 266071)

**摘要:**生物检索特别是按物索名因分类系统和标准迥异, 一直是分类学的难题, 迫切需要一套检索手段灵活、适用不同分类手段的快速检索、查询工具。为此, 作者在海岸带和海岛资源调查资料以及各位专家已有工作成果的基础上, 加强分类方法学的研究, 尝试开发了多媒体数据库和智能化的“海洋生物检索查询系统”(软著登字第 0012717 号), 探索利用数据库检索技术提高海洋生物检索查询效率、准确度的方法。

**关键词:** 渔业, 海洋生物, 查询, 检索系统, 分类

中图分类号: Q959; Q949; G257.36; G257.5 文献标识码: A 文章编号: 1000-3096 (2004)06-0018-05

## 1 研究背景

海洋是生命的摇篮, 海洋生物种类繁多, 海洋生物体中含有丰富的活性物质, 是人类食品和药物的资源宝库。随着与海洋生物有关的食物、生物制品、生物材料在国计民生中地位的提高, 海洋生物专业名词及数据已渗透到生命科学研究以及工农业、医疗卫生等生产实践的诸多领域。但是, 由于种种原因, 海洋生物学术资料与数据的数字化储存和数据共享应用一直未能得到充分重视和开发, 重复研究、低效率查询、甚至查不到资料的现象经常发生, 加上缺乏合理的数据处理系统, 数据资源的再加工和再利用能力低下, 为合作研究、教学和指导生产带来诸多不便。

我国有近万名从事与海洋生物有关专业的科研人员和科技工作者, 历年来在生物资源调查、分类学研究和产业应用研究等领域积累了大量资料和数据, 其中包括 20 世纪 50 至 60 年代的全国海洋普查和 80 年代的海岸带资源与环境调查、90 年代的海岛资源综合调查, 以及与国外发达国家合作开展过区域性的生物资源调查研究<sup>[1]</sup>。这些调查研究前后历时数十年, 前后耗资数千万, 倾注了几代海洋科学家的心血, 是不可多得的第一手科学资料。在工作实践中, 作者深刻体会到, 保护好已有的科研成果, 建立渔业生物资源的应用系统, 实现数据共享, 在技术上和内容上

跟踪国际前沿, 是科研、教育和生产所急需的, 其社会和经济效益难以估量。

山东省拥有全国三分之二的海洋生物资源研究力量, 海洋生物研究和数据资源储备有一定基础, 是开展建库工作的理想省份。目前, 国内已经有部分海洋生物的图鉴或名词等工具书出版发行, 部分解决了查询难的问题, 为计算机查询系统的建立打下了一定工作基础, 但资料更新需要再版印刷, 由于受发行量和经费等限制, 数据更新周期往往长达数年, 且存在查询不便、数据难以共享等弊端。建立计算机检索系统是解决上述问题的有效途径, 本文尝试从海洋生物专用名词入手, 以现有数据资源为基础, 从系统实用功能的设计角度, 探讨建立海洋生物共享数据库和高效查询系统的途径和模式。

## 2 设计目标

系统的主要设计目标是解决不同分类系统资料

收稿日期: 2003-07-20, 修回日期: 2004-04-06

作者简介: 杨俊杰 (1963-), 男, 山东青岛人, 副研究员, 电话: 0532-5818304, E-mail: yjj@smsta.com.cn

共享、再利用问题。

据统计,世界现存海洋生物 16 万多种,分属 49 个门,其中约 30% 是海洋特有的。按照学术、生产、市场等不同应用目的,每一个生物类群又可能有一个、两个甚至是多个分类体系,包括不同的生物特征描述。经过数百年的演化,形成了多种海洋生物分类系统,从 17 世纪的人为分类法、自然分类法,到 19 世纪以来的以比较形态学、解剖学、生理学、生物学、生物技术等现代生物分类学手段为基础形成的各种现代海洋生物分类系统,不同的分类方法各代表了不同历史时期科技发展的烙印,同时也积累了特征各异、丰富的分类经验和文献素材。

但是,不同的分类系统之间在分类依据、分类标准、沿用习惯等方面存在一定差异<sup>[2,3]</sup>,非专业研究人员使用单一的专业检索表往往不能得到所有的资料,“按名索图”和“按物索名”经常要翻阅大量资料,生物的鉴定往往必须依靠专家。目前,尚未查阅到有统一数据库建库标准的海洋生物分类查询检索工具。

本系统的设计目标,中心是围绕如何以国际公认海洋生物资源数据库的建库标准为参考<sup>[4]</sup>,建立适合不同分类系统的、适合我国国情的、实用的、智能化的渔业生物资源数据浏览、查询、分析和输出计算机系统模块。

经过对国内外海洋生物分类系统、海洋生物数据库进行比较,按照我国的学术分类研究基础和行业分类、部门划分和系统应用对象类型,系统具体的设计目标以满足以下要求时较为合理。

## 2.1 适用于不同的分类检索系统

建立智能化识别模块,兼容识别不同分类标准和依据的关键名词、概念、测量数据,利用数据挖掘技术原理使不同分类系统互为兼容。这是本系统设计的关键所在。

## 2.2 适应多种功能需求

通过将学术、生产、市场等不同应用方向的概念、名词、解释、图谱等设计成集成式的分类模块,实现用户可自由选择的分类标准化,以保持多领域不同用户群体的分类与查询习惯。

## 2.3 智能化用户模块

最大程度地满足用户检索、查询、比较分析、打印、保存查询结果、自动生成特定排序的速查手册等需求。并采用“用户应用表”技术和“智能查询器”技

术来满足学术、教学、生产、咨询等不同用户群体的查询需要。

## 2.4 开放式、交互式的多媒体数据库系统

为保证数据格式与国际接轨,数据库建模采用国际最流行的数据采集方法,按科学分类和市场分类原则进行分类;采用复嵌合思路和开放式设计构建数据库。用户可使用集浏览、查询与资料利用功能为一体的、WINDOWS 风格的数据浏览器,系统可自动为新品种建立数据库,用户也可以在已有数据库基础上全部或部分复制已有内容,生成自己的数据库,为高效工作提供方便。

## 2.5 快速查询

向导式实现用户查询条件、查询结果的保存、打印和查询条件的导入功能。快速并智能化的查询项识别、自由字段、组合查询功能,平均每千条记录的查询时间在 0.2 秒,中文记录在 3 万条以内时,查询时间不超过 20 秒。通过神经网络式的“触感”查询模块,用户可在数据库内容范围内所有数据表的字段进行单一或组合查询,也可进行以任意录入字为查询项的模糊查询;同时从任一数据项,可自由地将查看该数据所有相关的内容,也可进行同类数据的比较分析。

## 2.6 便于数据共享

用户可在任意窗口查询、储存、编辑相关的专业文章、档案。数据内容可由 ASFA 等系统中导入,也可由用户从自写文件中导入(.doc、.xls、.txt 等格式)。数据格式与国际专业数据库接轨,系统容易升级并便于网络动态更新。

# 3 数据库内容构成

根据设计目标,数据库的内容组成如下。

## 3.1 数据库内容结构<sup>[5,6]</sup>

数据库内容结构见表 1。不同海洋生物种类的数据库内容结构因应用和研究价值不同而存在较大差异,如浮游生物的数据项相对较少而开展研究较多的鱼类的数据项则较多。在上例中,部分图文字段项为基本库内容;生理生化有关的字段项为与国际接轨的内容,目前我国开展的研究尚少,建立此类数据项主要是为了储存发达国家的共享信息,第 10 项为用户可选内容。

## 3.2 数据类型

可以为短文本(名词概念类)、长文本(文献资

表 1 数据库内容结构

Tab. 1 Structure and fields of the database of halobios searcher

序号	数据类别	数据内容
1	名称	中、英、拉丁文通用名称、科学名称、商业名称、地区俗称;鱼类等经济生物包括 FDA(U. S. Food & Drug Administration)、SPRC(Seafood Products Research Center, FDA)的 RFE(Regulatory Fish Encyclopedia)等指定的名称标志,同名词和 WWW 页检索代码等
2	学术分类	流行分类方法的系统分类地位(界、门、纲、目、科、属、种、亚种),鉴定年代和鉴定人等信息
3	外貌特征	外观图片、体长、体重、形状、体色、特征构造等
4	生理特征	特征器官解剖结构示意图、生殖器官、生活史等
5	蛋白质特性	蛋白质分析图谱等
6	DNA 特性	DNA、基因分析图谱等
7	地理分布	地区(世界、中国)分布、生态(垂直、水平、温度、盐度)分布
8	经济用途	行业与产业用途;重要提取物及产品名称
9	毒性危害	生理毒性、危害性、危害对象等
10	参考文献	可在系统引导下由用户自己输入文件(论文、文章等)建库

料)、科学测量数据、分析数据、扫描图片、显微照片、音频和视频资料等等。

### 3.3 关键查询索引

任意单一数据项或组合数据项。

## 4 系统主要功能模块设计

系统设计总体上分为数据读取录入、用户系统和数据更新 3 个部分。(1) 数据读取功能将已有数据来源进行过滤并将有效数据转化为一定格式的文件,然后由用户系统自动读入数据库中,此项设计可大量节省已有数据的整理和处理的工作量。(2) 用户系统功能按设计目标采用第四代计算机语言设计而成,重点强化查询、比较分析和自动生成手册等智能化处理功能,突出对检索的速度要求,强调用户界面的简洁、易懂和使用方便。(3) 数据更新功能以 SQL 语言原理构建模块,挂接在用户系统中,同行用户之间可以互为数据供应者,用以数据交换和通讯和网上更新。

经过两年多的开发研究,科学数据检索查询系统的模块设计和数据库建模工作已基本完成。通过抽取历次海洋资源调查所取得数据和 SPRC 的 REF 数据页的取样数据,在系统中进行模拟运行,基本实现了设计目标的各种系统功能,证明设计思路完全可行。

## 5 系统的基本构成和基本功能

系统构成及功能流程见图 1。

系统由专业数据库和资料浏览分析器构成。数据库由基本数据库和用户数据库两部分组成,基本数据库可不断更新,用户数据库则可根据用户的具体需要进行编辑、修改。资料浏览分析器包括:(1) 系统登录模块:系统登录可由用户设定密码,限制系统登录的权限。(2) 资料浏览模块:为常见的资料查询模块,浏览文字、图片、数据等各类资料。(3) 智能化检索、查询模块:用于快速查找资料。(4) 组合、对比分析模块:根据各专业特点,引入科学分析原理,对检索出的资料进行分析,是科研分析的有力工具。

另外,系统设计了右键快捷菜单、数据更新模块、用户自定义手册模块和系统提示与帮助文件,分别用于快速浏览、网络更新资料、用户间数据交换共享、生成速查手册和打印输出,以及帮助文件指导用户正确使用系统。

## 6 实现信息共享

目前,系统的基本功能模块已经开发完毕,部分示范数据将陆续入库。数据库内容的不断补充和不断丰富,则需在有关部门的组织和支持下,依靠海洋生物学家、水产学家和水产科技工作者的共同努力完成。可以预期,日益发展的计算机多媒体和数据库技术将不断促进多媒体生物查询系统的研究开发,为专业文献资料和科研、资源数据的共享提供便利工具。

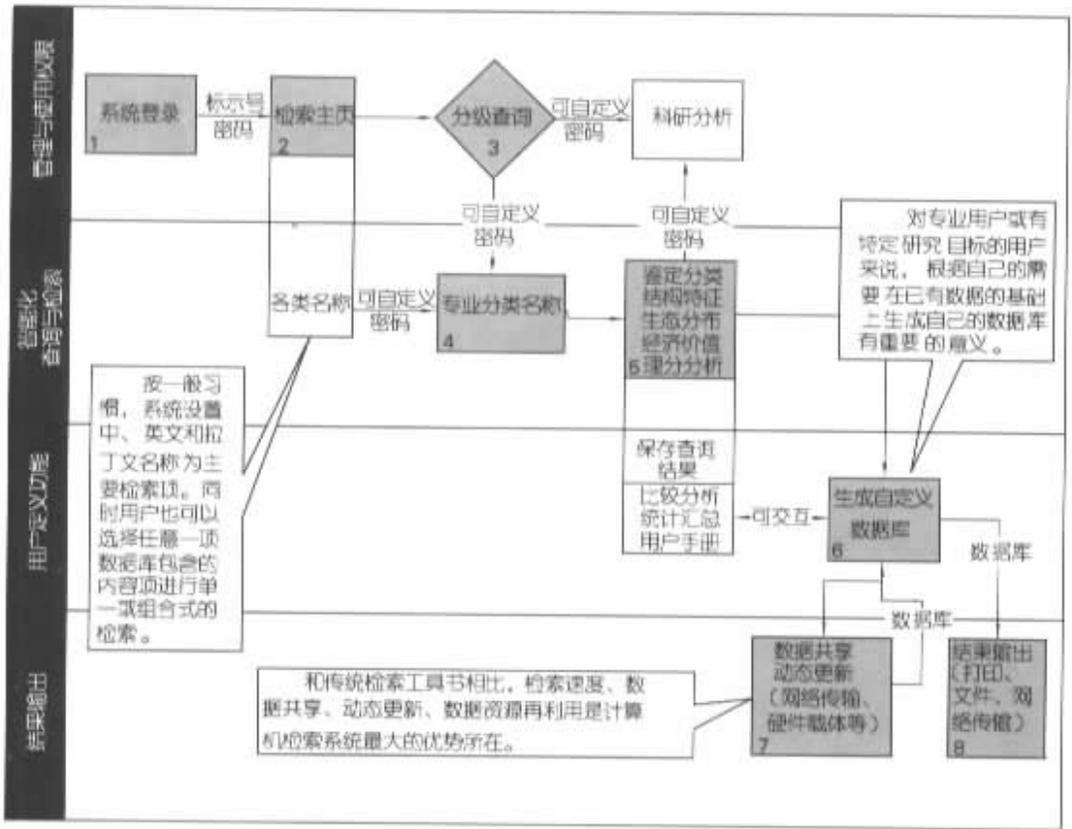


图 1 系统构成及功能流程示意

Fig.1 Structures and functions

参考文献：

[1] 山东省科学技术委员会. 山东省海岸带和海涂资源综合调查报告集综合调查报告[M]. 北京: 中国科学技术出版社, 1990. 269 - 458.

[2] 齐钟彦, 马绣同, 王祯瑞, 等. 黄渤海的软体动物[M]. 北京: 农业出版社, 1989. 265 - 276.

[3] 王如才. 中国水生贝类原色图谱[M]. 杭州: 浙江科学技术出版社, 1988. 247 - 255.

[4] 盖明举. 水科学和渔业情报系统数据库标引与用户指南[M]. 北京: 海洋出版社, 1990. 19 - 45.

[5] FISHBASE. FISHBASE[DB/OL]. <http://www.fishbase.org/> 2003 - 07 - 12/2003 - 07 - 05.

[6] FAO. List of fishery databases and applications [DB/OL]. <http://www.fao.org/fi/statist/fisoft/fisofoc.asp>, 2003 - 07 - 10/2003 - 06 - 30.

# Studies on techniques for function modules of halobios searcher

YANG Jun - jie<sup>1</sup>, HU Jian - ting<sup>1</sup>, LIU Bao - zhong<sup>2</sup>, GAI Ming - li<sup>1</sup>, SU Wen<sup>1</sup>, ZHOU Xian - ting<sup>1</sup>

( 1. Shandong Marine Engineering Exploration Academy, Qingdao, 266071, China; 2. Institute of Oceanology, the Chinese Academy of Sciences, Qingdao, 266071 )

**Received:** Jul., 20, 2003

**Key words:** fishery; marine organisms; search and inquire; computer system; taxonomy

**Abstract :** This paper introduces the development, structure and function of a novel taxonomy searcher " halobios searcher" [Marine Biology Taxonomy Searcher (National Software registration number: 0012717)] . Biology search and inquiry via organisms is complicated by various problems related to taxonomic systems, identification standards and requirement variances. The existence of an integrated identification tool would greatly improve the current situation for this purpose a multi - database system has been developed using data from the national marine resources survey and data from a number of other relevant sources this new search tool is expected to be useful to biologists, pharmacologists, aquaculture and fisheries environmental protection agencies and biomaterial manufactures.

(本文编辑 :刘珊珊)