

# 远洋渔业数据库设计和索引探究

王军<sup>1</sup>, 李阳东<sup>2,3,4</sup>, 李玉伟<sup>2,4</sup>

(1. 上海市海洋管理事务中心 上海 200050; 2. 上海海洋大学海洋科学学院 上海 201306; 3. 上海海洋大学大洋生物资源开发和利用上海市高校重点实验室 上海 201306; 4. 大洋渔业资源可持续开发教育部重点实验室 上海海洋大学 上海 201306)

**摘要:**我国远洋渔业已经走过30个年头,通过渔业生产、海区调查等方式积累了大量的远洋渔业数据。可以预见的是,随着时间的推移以及科学技术(如RS、GIS等技术)在远洋渔业方面的应用,远洋渔业数据必将具有海量特征。如何高效管理这些海量的远洋渔业数据是本研究要解决的关键。文章通过对远洋渔业中3种经济鱼(金枪鱼、竹荚鱼、鱿鱼)的生产、调查、地理等相关方面数据的分析,基于SQL Server 2000设计了远洋渔业调查数据库,基于Geodatabase设计了远洋渔业空间数据库。通过数据库的形式,实现了远洋渔业海量数据的高效管理。同时,针对所建立的远洋渔业数据库的空间数据建立了G树索引,为高效查询相关空间数据提供了支持。

**关键词:**远洋渔业;数据库设计;空间索引;渔业管理

中图分类号:F592.7;X321 文献标志码:A 文章编号:1005-9857(2016)01-0098-05

## On the Design and Index of Pelagic Fisheries Database

WANG Jun<sup>1</sup>, LI Yangdong<sup>2,3,4</sup>, LI Yuwei<sup>2,4</sup>

(1. Shanghai Administration Center for Ocean Affairs, Shanghai 200050 China; 2. College of Marine Sciences, Shanghai Fisheries University, Shanghai 201306 China; 3. The Key Laboratory of Shanghai education commission for Oceanic fisheries resources exploitation, College of Marine Sciences, Shanghai Fisheries University, Shanghai 201306 China; 4. Key Laboratory of Sustainable Exploitation of Oceanic Fisheries Resources (Shanghai Ocean University), Ministry of Education, Shanghai 201306 China)

**Abstract:** Lots of pelagic fishery data had been accumulated by fishing production and ocean surveys during the passed 30 years development of pelagic fisheries of our country. It could be predicted that with time going on and science and technology (such as RS, GIS and other technologies) applying in pelagic fisheries, there would be a flood of data. It is the key point that how to effectively manage these vast quantities of pelagic fishery data in this paper. The data of three economic fish species (tuna, mackerel and squid) had been analyzed, which included production, research, geography, and other related data. Survey database of pelagic fishery had been designed based on the SQL Server 2000 and spatial database had been designed based on Geodatabase. Efficient management of huge amount of data was implemented by this form of database. Simultaneously, a G-tree index for this database of pelagic fishery was established to query data quickly and efficiently.

**Key words:** Pelagic fisheries, Design of database, Spatial index, Fishery management

远洋渔业作为海洋经济的重要组成部分,已经涉及国家的海洋权益、经济利益、食物供应安全等方面,我国对远洋渔业给予了高度的重视。同时,海洋渔业管理是海洋管理的重要内容,做好渔业管理,对海洋渔业资源开发与管理起到重要的积极作用。

自1985年来,我国海洋工作者在进行远洋捕捞的同时,对多数海区进行了大规模的调查,积累了大量的远洋渔业数据。他们对这些数据进行分析处理,独立开发建设了很多数据库。毕健等<sup>[1]</sup>利用来自于舟山、上海、烟台、大连、宁波等17个渔业公司(或单位)提供的1994—1995年渔获量统计资料建立了西北太平洋鱿鱼钓数据库,利用Foxpro数据库系统准确快速处理鱿鱼钓渔获量数据,利用经济领域中的预测方法来预测产量、可投入船数、平均日产量等,指导渔业生产与管理。李小恕等<sup>[2]</sup>利用1997—2001年对海洋生物资源与生物栖息环境调查数据,建立了海洋生物资源与生物栖息环境数据库,该数据库在海洋渔业资源与环境的研究、评价中得到广泛应用。袁骥等<sup>[3]</sup>利用中国水产科学院黄海、东海和南海水产研究所1999—2002年多个航次的调查数据,设计了海洋渔业生态环境监测数据库系统,针对海洋渔业生态环境监测特殊需要,增加了污染物质生物体残留量子数据库以及多种水质、底质和生物污染程度综合评估模型,为海洋渔业生态环境质量评价和海洋生态环境保护提供相关数据和信息。陈卫忠等<sup>[4]</sup>在系统收集东海区历年海洋捕捞产量、捕捞努力量统计资料、主要经济鱼类生物学参数以及渔业资源研究文献报告等资料基础上,对数据资料进行整理和补充,建立了渔业资源研究数据库。张寒野<sup>[5]</sup>在对海洋生物资料调查信息分析与整理的基础上,将涉及的数据分为属性数据和空间数据两种,分别建立了海洋生物资源数据库和矢量图层;最终利用这些数据库建立了海洋生物资源信息管理系统,通过该系统用来准确评估海洋渔业资源现状。

总体上,目前这些与渔业相关的数据库存在以下不足:①数据没有统一规范的格式,导致处理数据时费时费力,增加了操作失误和人为误差<sup>[6]</sup>;②数据库中数据的丰富度、完整度不够<sup>[7]</sup>,很少将资源、

环境、空间、人文、科技等相关信息关联起来,涵盖的信息量不足,导致数据库水平不高;③数据库相对孤立、分散,大多针对单一用户提供单一产品,应用面较窄,共享性差<sup>[8]</sup>,数据库之间难以实现互联互通,信息闭塞、推广难度大<sup>[9]</sup>;④数据库比较小型化,没能形成综合的大型数据库,给远洋渔业数据的科学规范管理带来巨大挑战。

依据现有的数据库且结合上海海洋大学海洋渔业遥感GIS技术实验室的数据,兼顾当前RS以及GIS技术获得资料,拟对远洋渔业数据库进行设计开发,设计一个规范、全面、实用性强的远洋渔业数据库。并利用该数据库对这些多源异构的海量远洋渔业信息进行管理,以便为海洋渔业科研奠定更好的基础,为海洋渔业生产提供多面服务,为渔业资源合理利用提供科学指导,为海洋资源管理做出科学决策。

## 1 远洋渔业数据库设计

在前人的基础上,对远洋渔业涉及生产、环境、资源、地理、船舶等数据进行了分析和梳理,将涉及的数据分为调查数据和基础数据两种形式,分别设计了远洋渔业调查数据库和基础数据库(即空间数据库)。

远洋渔业调查数据库在SQL Server 2000的环境下,根据需求创建的调查信息数据库。该数据库本着信息规范、系统性能稳定、数据安全、信息完整、系统可扩展等原则,创建了包括鱿鱼、金枪鱼、竹荚鱼生产信息和生产调查生物学数据信息专题,以及生产调查环境信息专题等的11个数据库表以及若干字典表。

远洋渔业地理空间数据库在ArcGIS的环境下,基于Geodatabase根据实际渔业调查或者生产中涉及的数据创建的空间数据库。该数据库包括经纬网、FAO渔区、海区、捕捞点等具有特定地理意义的信息。这些包含地理信息的空间数据大多以矢量数据格式存在,通常采用“图层”来对它们进行组织和管理,各个图层分别存储不同专题的空间信息。在本数据库中,包含捕捞点、观测点、航线、渔区、海区、洲等6个含有基础空间数据的图层。

### 1.1 远洋渔业数据库需求分析

远洋渔业数据库的数据信息包含调查数据和

空间基础数据,随着今后调查和渔业生产活动的执行,积累的数据会不断增长。该数据库的调查数据包括与调查相关的数据,如调查时间、调查经纬度、网次、温度、深度、盐度、渔获物量以及渔获物生物学方面的数据;基础空间数据主要指海洋功能区划信息这类代表特定地理意义的数据,包括经纬线、渔区、海区、海陆边界、观测点、作业地点等。这些数据在数据库中以图层的形式出现,同时调查数据中涉及空间的数据(如经度、纬度)均属于空间数据。该数据库中的调查数据(是指基本上与空间位置没有直接关系的数据)存放于 SQL Server 数据库中。

## 1.2 远洋渔业调查数据库建立

### 1.2.1 远洋渔业调查数据库概念结构设计

概念结构设计是将分析得到的用户需求抽象为概念模型的过程。即在需求分析基础上,设计出能够满足用户需求的各种实体以及它们之间的相互关系概念设计模型。概念模型是对信息世界建模,能够方便、准确地表示信息世界中的常用概念<sup>[10]</sup>。概念结构设计能真实、充分地反映现实世界及其事物与事物之间的联系,易于理解和更改,以及易于向关系、网状、层次等其他数据模型转换。

在远洋渔业数据库中,每艘船舶会在不同的渔区从事不同的作业,这些作业包括对不同经济鱼种进行生产以及生产时的环境调查等。在生产的过程中,对捕获物进行统计和测定。不同的捕获种类有着不同的生物学特征,并且不同的经济鱼种在生产过程中有着不同的生产方式。所以,在本数据库中涉及对象有:金枪鱼、鱿鱼、竹荚鱼生产信息,环境调查信息,生物种类,生物学信息,海区信息,船舶信息等。用 E-R 图表示它们之间的关系如图 1 所示。

### 1.2.2 远洋渔业调查数据库逻辑结构设计

E-R 模型是用户的模型,它独立于任何一种数据模型。因此,需要将用 E-R 图表示的概念模型转换为某个具体的数据库管理系统所支持的数据模型,然后建立用户需要的数据库。

根据数据库概念结构设计阶段得到的远洋渔业地理属性数据库 E-R 模型,我们可以设计以下

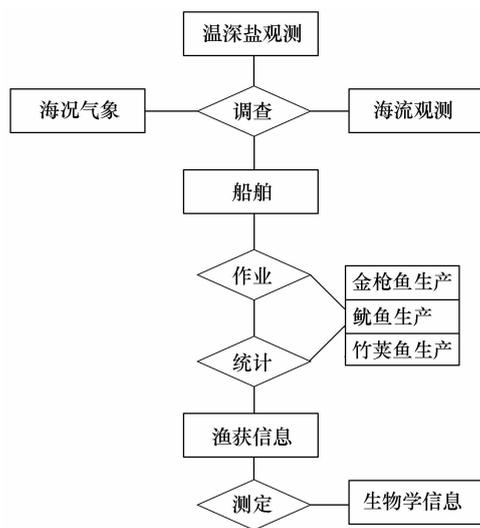


图 1 远洋渔业数据库 E-R 图

注:长方体表示实体,菱形表示实体间的关系。

远洋渔业调查数据库逻辑结构模型,其中实体标志码用 \* 标出(下同)。

(1) 鱿鱼生产信息,包括:日期\*, 渔业公司\*, 经度, 纬度, 作业渔船数, 渔获量, 作业类型, 备注。

(2) 鱿鱼生产调查生物学信息,包括:日期\*, 船名\*, 渔业公司, 船舶呼号, 经度, 纬度, 胴长, 体重, 性别, 性成熟度, 摄食等级, 胃含物, 备注。

(3) 金枪鱼生产信息,包括:作业日期\*, 船名\*, 船舶状态, 投绳船速, 投钩起点时间, 投钩起点经度, 投钩起点纬度, 投钩终点时间, 投钩终点经度, 投钩终点纬度, 起钩起点时间, 起钩起点经度, 起钩起点纬度, 起钩终点时间, 起钩终点经度, 起钩终点纬度, 两浮子间钩钩数量, 投钩数量, 投绳长度, 支绳长度, 两支绳间主绳长度, 投放主绳长度, 鱼种, 渔获尾数, 加工重量, 兼补海龟状况, 备注。

(4) 金枪鱼生产调查生物学信息,包括:日期\*, 船名\*, 渔业公司, 船舶呼号, 经度, 纬度, 胴长, 体重, 性别, 性成熟度, 摄食等级, 胃含物, 备注。

(5) 金枪鱼生物学统计信息,包括:日期\*, 船名\*, 渔业公司, 船舶呼号, 经度, 纬度, 渔获种类, 平均体长, 平均体重, 备注。

(6) 竹荚鱼生产信息,包括:日期\*, 船名\*, 网次\*, 渔业公司, 船舶呼号, 放网时间, 放网纬度, 放网经度, 起网时间, 起网经度, 起网纬度, 产量,

CPUE,拖速,拖向,网位深度,网口高度,网型,手纲,沉力,浮力,水平扩张,曳绳长度,鱼群下缘,鱼群上缘,鱼群水深,鱼群高度,鱼群影像,网位仪水深,备注。

(7)竹荚鱼生产调查生物学信息,包括:日期\*,船名\*,渔业公司,船舶呼号,经度,纬度,叉长,体重,纯体重,体宽,体高,体周,性别,性成熟度,摄食等级,胃含物,备注。

(8)海况气象,包括:日期\*,船名\*,时间\*,渔业公司,船舶呼号,经度,纬度,天气,气温,气压,流速,流向,风速,风向,能见度,浪高,浪向,干球温度,湿球温度,总云量,低云量,备注。

(9)温盐深观测,包括:日期\*,船名\*,时间\*,渔业公司,船舶呼号,经度,纬度,深度,温度,压力,密度,盐度,叶绿素-a,溶解氧,备注。

(10)海流观测,包括:日期\*,船名\*,时间\*,渔业公司,船舶呼号,经度,纬度,水深,速度 $x$ ,速度 $y$ ,速度 $z$ ,温度,速度,方向,备注。

(11)船舶档案,包括:船名\*,船舶呼号\*,所属公司,船籍港,船籍国,船型,作业许可证,全长,功率,总登记吨位,巡航能力,造船时间,冷冻能力,仓容量,作业方式,船舶照片,备注。

(12)渔区:包括:渔区 ID\*,渔区名称。

(13)海区:海区 ID\*,海区中文名称,渔区英文名称。

### 1.2.3 远洋渔业地理数据库物理结构设计

数据库在实际的物理设备上的存储结构和存取方法称为数据库的物理结构,与给定的硬件环境和 DBMS 软件产品有关。本数据库是在 Windows XP 的操作系统下,磁盘 320 G 的环境中,利用 SQL Server 2000 建立。

### 1.3 远洋渔业空间数据库设计

本数据库是基于 Geodatabase 建立。在 Geodatabase 中,要素类是具有相同的属性集、相同的行为和规则的空间对象的集合。所有的数据都在同一数据库中存储并中心化管理,实现地理数据的统一存储管理。同时,还可以实现无缝、无分块的海量要素的存储<sup>[11]</sup>。

根据空间数据的特征,对空间对象建立的逻辑

结构如下:

航线信息,包括:要素 ID\*,形状,要素类型,航行航次,航行船名,开始航行时间,终止航行时间,航行海区。海区信息,包括:要素 ID\*,要素类型,陆地中文名,陆地英文名。渔区信息,包括:要素 ID\*,形状,要素类型,渔区号,渔区所属海区,面积。洲信息,包括:要素 ID\*,形状,要素类型,洲中文名,洲英文名,面积。影像信息,包括:影像名称,影像 ID\*,存储路径,文件格式,空间分辨率,影像格式,拍摄时间,左上角经度,左上角纬度,右下角经度,右下角纬度。

根据空间数据的逻辑结构,建立空间数据库的物理结构如表 1 至表 3 所示。

表 1 航线和海区信息

航线信息		海区信息	
字段名称	字段类型	字段名称	字段类型
要素 ID	int	要素 ID	int
形状	geometry	形状	geometry
要素类型	char	要素类型	char
航行航次	int	海区号	int
航行船名	char	海区中文名	char
开始航行时间	date	海区英文名	char
终止航行时间	date	面积	double
航行海区	char		

表 2 渔区和洲信息

渔区信息		洲信息	
字段名称	字段类型	字段名称	字段类型
要素 ID	int	要素 ID	int
形状	geometry	形状	geometry
要素类型	char	要素类型	char
渔区号	int	洲中文名	char
渔区所属海区	char	洲英文名	char
面积	double	面积	double

表 3 影像信息

字段名称	字段代码	字段类型	字段长度
影像名称	L_Name	char	10
影像 ID	ID	int	4
文件格式	shape	char	10
空间分辨率	S_Resolution	float	4
影像格式	Type	char	10
拍摄时间	T_Time	datetime	4

续表

字段名称	字段代码	字段类型	字段长度
左上角经度	L_Lon	double	4
左上角纬度	L_Lat	double	4
右下角经度	R_Lon	double	4
右下角纬度	R_Lat	double	4

针对不同的图层,在方便管理和储存的前提下,利用 Geodatabase 建立了远洋渔业空间数据库。在该数据库中,要素类型共分为 3 种:点状、线状和面状,共有 7 个图层。捕捞点、观测点属点特征图层;航线属线特征图层;渔区、海区、经纬线、洲等属于面特征图层。在 Geodatabase 中的个人数据库中,利用 ArcGIS Catalog 工具,将盐度、温度和叶绿素的影像存储在该数据库中。

## 2 远洋渔业数据库索引的建立

渔业调查数据库中涉及大量的空间调查数据,这些数据是具有点实体特征,并且均与时间有关。实际应用中,经常需要对这些渔业数据进行空间和时间的查询访问。为了快速而又准确的查找到所需目标,作者根据远洋渔业数据库中的调查时间数据建立了顺序索引,对空间数据库建立了 G 树索引。

由于调查数据都会有时间的字段记录,因此可以针对时间建立顺序索引。顺序索引建立的方法有 B-树、B<sup>+</sup>-树索引等。这样,要查找某一时间的调查数据,按照顺序检索或者折半检索,就可以快速定位满足条件的数据。

G 树格网索引将海区作为第一层次,渔区作为第二层次,点状实体、线状实体、多边形实体等作为第三层次。

在检索过程中,先对数据进行过滤:先检索第一层次,找到目标实体所在的区;将此层次中目标实体所在的海区的格网投影到第二层次的格网中,找出目标实体所在的渔区;再将第二层次中目标实体所在的渔区格网投影到第三层次中,然后在第三层次中进行精确查找,直到精确查找出该目标实体。

## 3 结论

本研究以现有的远洋渔业生产数据、调查数据

以及地理数据为基础,在详尽分析了远洋渔业中金枪鱼、竹荚鱼和鱿鱼等数据资料之后,依据分析结果进行了远洋渔业数据库的设计与探究。本数据库分为两部分,利用 SQL Server 2000 建立了远洋渔业属性数据库,利用 ArcGIS 建立了远洋渔业空间数据库,使得基础数据与空间数据得以连接在一起。另外,还对远洋渔业空间数据库的索引进行了研究并针对该数据库建立了 G 树索引,提高了空间数据查询的效率。远洋渔业数据库的建立,为海洋渔业科研奠定更好的基础,为海洋渔业生产提供多面服务,为渔业资源合理利用提供科学指导,为海洋资源管理做出科学决策。

该远洋渔业数据库设计目前仅考虑金枪鱼、竹荚鱼、鱿鱼 3 种经济鱼,下阶段的工作是把其他经济鱼种也考虑进来,同时在该数据库的基础上开发远洋渔业数据库系统,实现远洋渔业有关的分析和统计等功能。

## 参考文献

- [1] 毕健,陈新军.西北太平洋鱿鱼钓数据库的开发和应用[N].上海水产大学学报,1996,5(3):205-208.
- [2] 李小恕,李继龙,贾静,等.数据库技术在海洋生物资源与生物栖息环境数据管理中的应用[J].上海水产大学学报,2002,11(4):371-377.
- [3] 袁骥,沈新强.海洋渔业生态环境监测评估数据库的设计和实现[C]//中国水产学会 2003 年学术年会,2003.
- [4] 陈卫忠,李长松,胡芬,等.东海区海洋渔业资源研究数据库系统的设计和实现[J].中国水产科学,2001,7(4):91-94.
- [5] 张寒野.东海海洋生物资源信息管理系统开发[D].上海:上海水产大学,2006:1.
- [6] 黄巧珠,吕俊霖,麦丽芳.我国渔业科学数据库的现状与发展趋势[J].安徽农业科学,2009(5):15977-15978,16127.
- [7] 杜云虎,周成虎,邵全琴,等.海洋渔业数据库质量控制研究[J].中国图像图形学报,2002,7(3):276-281.
- [8] 杨宁生.现阶段我国渔业信息化存在的问题及今后的发展重点[J].中国渔业经济,2005(2):15-17.
- [9] 岳昊,欧阳海鹰,曾首英,等.浅谈中国鱼类多样性数据库建设现状:以 FishBase 为例[J].中国农业通报,2013,29(8):59-63.
- [10] 王珊,李胜恩.数据库基础与应用[M].北京:人民邮电出版社,2002:1-6.
- [11] 王新生,王红,朱超平. ArcGIS 软件操作与指南[M].北京:科学出版社,2010:25-27.