

学士学位论文

中期检查报告

|  |  |
| --- | --- |
| 论 文 名 称： | **船舶监测数据处理系统的设计与实现** |
| 姓 名： | **邓经纬** |
| 学 号： | **14005014** |
| 学院指导老师： | **马平** |
| 企业指导老师： |  |
| 企业老师单位： |  |

北京航空航天大学软件学院

2018年 4 月 8 日

目录

[1. 课题背景介绍 1](#_Toc5022481)

[1.1 课题来源 1](#_Toc5022482)

[1.2 课题目标与任务 1](#_Toc5022483)

[1.3 主要技术与成果 1](#_Toc5022484)

[2. 课题任务进展 2](#_Toc5022485)

[2.1. 需求分析 2](#_Toc5022486)

[2.1.1. 船舶数据调研 2](#_Toc5022487)

[2.1.2. 系统需求分析 2](#_Toc5022488)

[2.2. 系统设计 4](#_Toc5022489)

[2.2.1. 系统功能结构 4](#_Toc5022490)

[2.2.2. 系统总体设计 5](#_Toc5022491)

[2.2.3. 系统数据库设计 5](#_Toc5022492)

[2.3. 主要工作内容和完成情况 6](#_Toc5022493)

[2.3.1. 数据库设计和搭建 6](#_Toc5022494)

[2.3.2. 爬虫代码编写 6](#_Toc5022495)

[2.3.3. Node.js环境搭建和Express框架使用 7](#_Toc5022496)

[2.3.4. 高德地图API学习和使用 8](#_Toc5022497)

[2.3.5. Spark环境和Scala环境的搭建 9](#_Toc5022498)

[2.3.6. 服务器配置和搭建 9](#_Toc5022499)

[2.4. 总体任务完成情况 9](#_Toc5022500)

[3. 后期拟完成的研究工作及进度安排 10](#_Toc5022501)

[4. 参考文献 11](#_Toc5022502)

# 课题背景介绍

## 课题来源

来源于指导教师马平老师实验室的实际项目，在此基础上略作修改

## 课题目标与任务

本课题的目标是通过设计并实现一个船舶数据处理系统来实现船舶数据的存储、处理和分析。为查看船舶数据信息的用户提供可视化接口。也在一定程度上对数据进行尽可能的数据分析和挖掘发掘

本课题旨在解决一下几个关键问题：

1. 将已经收集到的大量数据进行有效，安全的存储。
2. 对数据进行进一步优化处理，挖掘。
3. 对挖掘成果和数据的可视化展现。

## 主要技术与成果

首先使用了Hadoop分布式存储和Mongodb数据库存储数据；对于船舶的设备数据主要使用将MySQL数据库导入Mongodb数据库，而船舶运行状态数据则使用爬虫技术获得并存入数据库；在数据挖掘方面，则使用Spark计算框架库编写代码计算获得价值信息，最终根据实际需求确定存入数据库或是直接展示给用户；可视化问则使用Node.JS、Express框架搭建网页、使用高德地图API在地图上展现各种位置和挖掘成果，主要包括船舶分布信息、单艘船舶航行热点、单艘船舶设备数据分析结果、速度分析结果等；使用ECHART框架构建数据图表，显示数据和挖掘结果。最终用户通过网页浏览信息，搜索查看船舶数据挖掘成果。

# 课题任务进展

## 需求分析

对于系统的用户，需要能够通过浏览器查看所有船舶的位置，查看船舶信息，查看所有船舶的分布热点；搜索某船舶的轨迹，搜索某船舶的航行热点，搜索某船舶速度分析结果；查看船舶设备数据及分析结果，查看船舶储物量数据和分析结果。这样就可以分析得出如图2.1的系统用户用例图。

对于船舶设备数据主要来源于船舶设备数据采集人员，需要能够将一定格式的数据文件上传导入数据库进行存储。

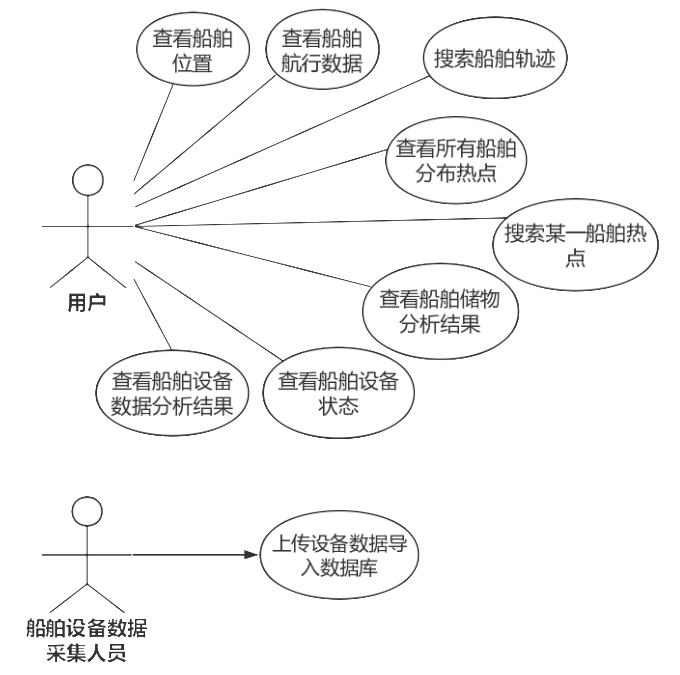


图2.1 系统用例图

在用例图的基础上进一步分析，得出了船舶设备数据采集人员业务流程图和用户业务流程图，分别图图2.2和图2.3所示。

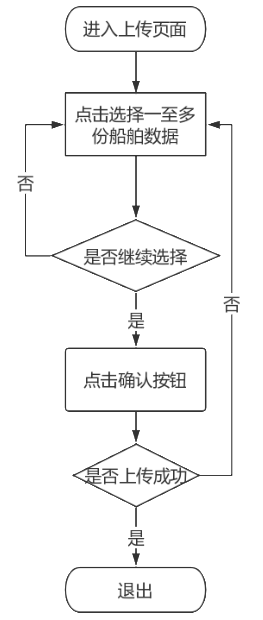


图2.3船舶设备数据采集人员业务流程图

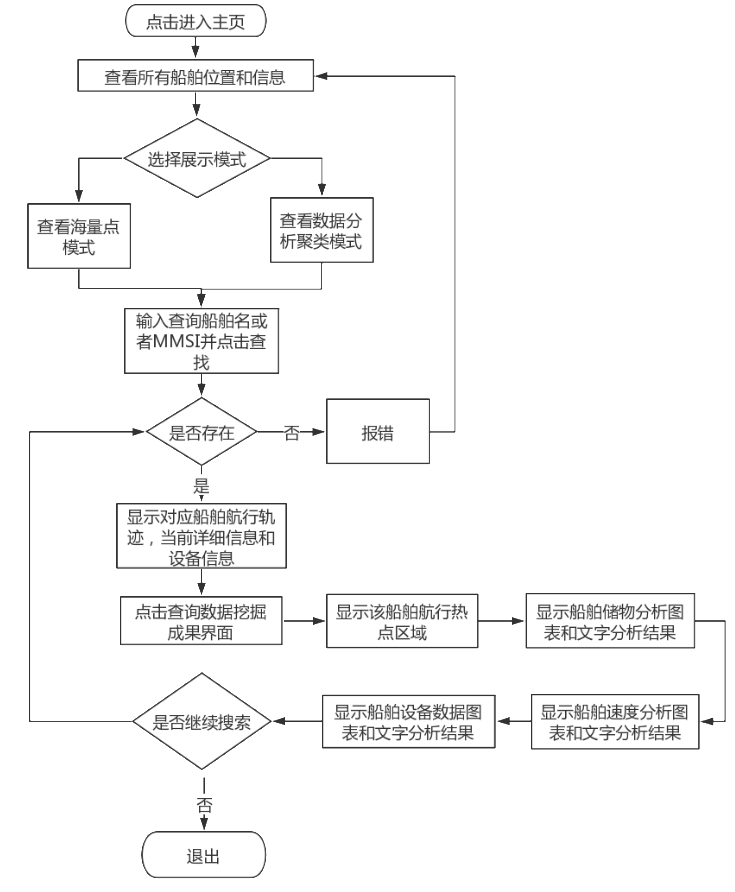


图2.4用户业务流程图

## 系统设计

### 系统功能结构

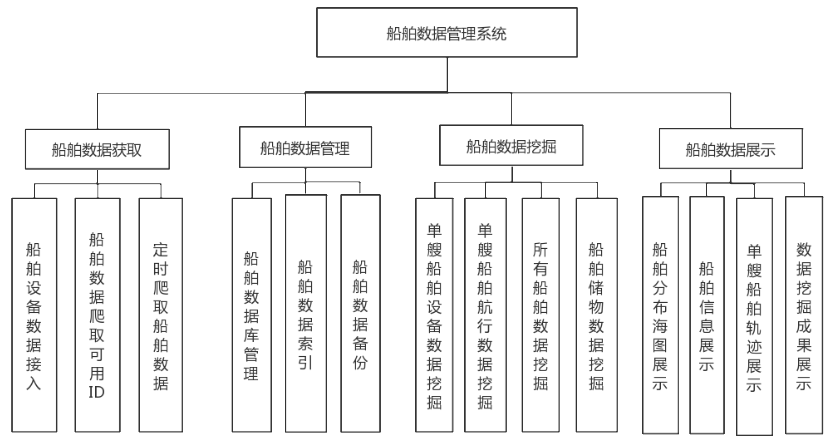


图2.5系统功能结构图

### 系统总体设计

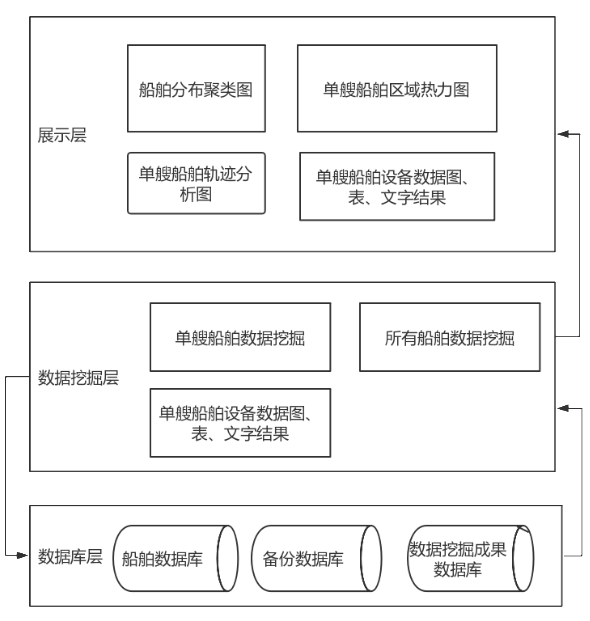


图2.3系统总体设计图

### 系统数据库设计

如图2.3所示（传统ER图不便于表述MongoDB的结构），数据库的详细设计分为五大部分[13]。首先shipinfo主要存储MMSI(水上移动通信业务标识码)、IMO（国际海事组织的识别码）等基本不会变动的信息；而shipdyna主要存储船舶位置（GPS）、船首向（朝向）、船舶状态（航行/系泊）等变动信息。而且考虑到在查询等功能场景中经常需要以名称、MMSI等字段查询。为了便于操作，添加了相应的字段；shipdynahis，首先是考虑到对于数据挖掘的任务要求设计的数据库，其字段绝大部分shipdyna相同，又因为所有动态信息历史纪录均对应一艘船舶，所以添加了shipinfo来建立对应关系。

对于船舶设备，可以获取的数据主要有发电机功率，主机平均转数，增压器转速，气缸排气量，气缸冷却水温度，燃料油量，柴油量，活塞冷却水温度，机油温度，减速箱油压，增压器压强，海水温度，机舱温度。同理也需要与对应的船舶建立关系。



图2.4数据库详细设计图

## 主要工作内容和完成情况

### 数据库设计和搭建

数据库详细设计已在2.2.3小结叙述。再次不做赘述。

其中需要重点强调的就是对于地理索引的建立和使用[13]。在建立地理位置索引时，就需要将经纬度以数组的形式存到GPS 字段中。使用命令行即可建立地理索引。在查找某条船最近的船舶，或者搜索一定范围的船只都可以使用参考文献[13]的方法进行查询。

此外还进行了数据库备份确保数据库安全。主要包括服务器快照，编写脚本代码自动备份，本地手动备份。

### 爬虫代码编写

经过调研分析发现几乎所有网站都暴露了对应的数据接口，而且经过对比分析发现所有数据几乎没有太大的差距。所以最终决定直接编写原生的爬虫代码对中国港口网(www.chinaport.com)进行数据爬取。

在按照系统功能结构设计的要求下，先爬取可用ID存入id.txt，在每隔2个小时读取ID爬取数据。之后对数据进行初步清洗，删除重复数据。

爬取过程中遇到了一些问题，并最终找到了解决方案。如表2.2所示。

**表2.2爬虫设计问题及解决方案表**

|  |  |  |  |
| --- | --- | --- | --- |
| 问题编号 | 问题名称 | 问题描述 | 解决方案 |
| 1.1 | 重复船舶 | 对于不同的ID，得到的船舶数据相同 | 发现网站ID每7个连续的ID对应相同的船舶。对ID每次加7 |
| 1.2 | 数据为空 | 对于一部分ID,得到全空的数据或者部分字段为空的数据 | 如果数据为空则舍弃；如果数据大部本字段为空，则统计空字段的数目。小于一定的阈值则舍弃 |
| 2.1 | 主机连接断开 | 当爬取一定数量的ID时，会因与主机的连接中断而停止爬取数据 | 添加不同的User\_Agent和代理IP池,每次爬取选择随机的选项进行爬取 |
| 2.1 | 多进程错误 | 当同时运行多个进程增加爬取效率会断开连接 | 同上 |

还存在的问题主要是爬取的船舶数量太少，时间较短，还不足以支持之后的数据挖掘操作。

### 服务器配置和搭建

这一部分主要包括Hadoop环境搭建，Hadoop与MongoDB连接，Spark和Scala环境搭建，绑定域名，使用Screen在服务器后端运行，配置Hadoop环境。Hadoop环境搭建，与MongoDB连接主要分为以下几步：

添加主机IP与域名的对应关系，相互导入密钥实现SSH免密登录。下载Hadoop和JDK安装包，配置HADOOP\_HOME和JAVA\_HOME等相关环境。在Hadoop对应文件夹根据主从服务器域名、个数和端口新建文件夹修改相应的配置文件。初始化Hadoop，运行测试程序测试。最终如图2.5所示



图5.2 Hadoop启动成功示意图

Hadoop和MongoDB连接，Spark环境配置基本类似，不做详细描述。

### 船舶设备数据的模拟和导入

由于无法获得大量船舶设备数据满足大数据和数据挖掘的需要。于是编写脚本每五分钟生成大量的船舶设备数据。并设计网页上传文件至服务器，通过服务器脚本每分钟读取数据，导入数据库。

### Node.js环境搭建和Express框架使用

**表2.3 Node.js环境和Express框架使用步骤**

|  |  |  |
| --- | --- | --- |
| 步骤编号 | 步骤名称 | 步骤详细描述 |
| 1.1 | 学习编写练习项目 | 图书馆借取两本Node.js实战教程[16]，在网络课程的帮助下编写了练手项目 |
| 2.1 | 新建网页项目并连接Mongodb数据库 | 利用mongoose [14]进行对象建模。首先配置项目中mongoose环境连接数据库 |
| 2.2 | 进行数据建模 | 建立与数据库相对应的Schemas以及相应的model来完成数据库与对象的相互映射 |
| 2.3 | 编写后端代码接口 | 针对不同的操作编写不同的数据库操作接口 |

所以主要编写了了在数据库各种查询操作的接口，其中也利用了Mongodb的地理位置索引方便查询。

目前已编写的接口主要有查询所有船舶位置和朝向，根据名字和MMSI查询特定船舶的详细信息，查询某一区域的船只，查询某船舶附近的所有船舶，查询某船舶的历史位置，搜索船舶设备信息，根据船舶名称查找MMSI。

之后还需要根据数据挖掘结果和需要编写其他数据库接口。

### 高德地图API学习和使用

主要通过高德地图的官方API文档[9]，最终将数据可视化展示在地图上。

目前已经完成的页面有所有船舶位置界面，所有船舶详细信息显示，某船舶的历史轨迹和回放，所有船舶聚合位置，所有船舶海量位置显示，船舶设备数据展示。还需要根据具体情况进行更多的优化。

还待开发的页面主要有船舶设备数据图表化，某船舶航行热点界面，以及数据挖掘分析得出的结果展示页面。

## 总体任务完成情况

表2.4 计划进度表

|  |  |  |
| --- | --- | --- |
| 开题计划时间 | 工作任务 | 完成情况说明 |
| 2018.12.27-2019.1.15 | 具体研究相关技术并初步实施设计 | 基本掌握 |
| 2019.2.1-2019.2.28 | 完善设计并实现数据库和接口 | 数据库相关基本实现，接口完成了一部分。另外根据之后具体需要编写 |
| 2019.3.1-2019.3.10 | 获取数据并将数据准确安全地存储 | 爬虫和数据库运行稳定，数据库安全备份。 |

# 后期拟完成的研究工作及进度安排

表2.5 后期任务计划表

|  |  |  |
| --- | --- | --- |
| 计划时间 | 工作任务 | 完成计划（技术/手段/方法） |
| 2019.4.1-2019.4.10 | 对可能的数据挖掘方向进行研究和可行性分析 | 1. 已有的参考文献和网络调研 2. Spark框架的学习 |
| 2019.4.11-2019.4.20 | 编写合适的接口，编写数据挖掘算法代码，得出相应的结果存入数据库或者直接展示 | 1. Spark计算框架[15] 2. Scala项目 3. 参考文献学习 |
| 2019.4.21-2019.5.1 | 设计网页，将数据挖掘成果可视化展现 | 1. 高德地图API [9] 2. Echart框架[10] |
| 2019.5.1-2019.5.10 | 网页界面组合，优化网页结构和显示结果 | 1. 代码优化 2. 网页结构调整 |
| 2019.5.10-2019.5.13 | 总结课题撰写毕业论文 | 1. Word文档 2. 毕业论文格式工具 3. 查重工具 |

# 参考文献

[1]谢芳.基于互联网的大数据挖掘关键技术分析[J].信息记.

[2]陈意惠,蒲小莲.基于PLC的船舶机舱实时监测报警系统[J].上海海事大学学报,2005,26(2):67-70.

[3]李永攀,刘正江,郑中义.基于时空密度的船载AIS数据聚类分析方法研究[J].重庆交通大学学报（自然科学版）,2018,37(10):117-122.

[4]钱晓江,宿巧丽.船舶航海信息采集系统(MIGS)的设计与实现[J].微型电脑应用,

2003,19(8):15-19.

[5]陈玺,马修军,吕欣.Hadoop生态体系安全框架综述[J].信息安全研究,2016,2(8):684-698.

[6]陈丽,黄晋,王锐.Hadoop大数据平台安全问题和解决方案的综述[J].计算机系统应用,2018,27(1):1-9.

[7]李建江,陈玮,李明, 等.基于网格热度值的船舶规律路径提取算法[J].计算机研究与发展,2018,55(5):908-919.

[8]徐思路. 基于船舶轨迹大数据的交通环境时空分析研究[D].大连海事大学,2018.

[9]向玉云,高爽,陈云红,黄嘉成,许新华.百度、高德及Google地图API比较研究[J].软件导刊,2017,16(09):19-21+25.

[10]丁东辉. 基于Hadoop和D3.JS的互联网+博物馆可视化平台的研究与实现[D].东华大学,2017.

[11]李晓英,何首武,苏树海.基于NoSQL的海量出租车GPS数据Web服务研究[J].广西科学院学报,2016,32(03):231-236.

[12]高立佳. 基于MongoDB与WebGL的船舶大数据可视化研究[D].大连海事大学,2016.

[13]宋平亮. 基于MongoDB的航道数据Web服务研究[D].大连海事大学,2015.

[14]郑睿博,潘雅静.基于Mongodb以及GIS的海量探地雷达数据存取研究[J].城市勘测,

2018(S1):277-280.

[15]周尧,刘超,徐树楠,曹振宇,耿丽丽,刘建川,廖一铧.基于Spark与MongoDB的地理空间大数据应用分析系统设计与实现[J].测绘与空间地理信息,2018,41(09):71-74.