

青海湖野生鸟类数据统计分析系统的设计与实现

万孙华，黄志一，周园春，阎保平
中国科学院计算机网络信息中心，北京 100190

摘要：青海湖自然保护区生活着大量的野生鸟类，保护区的工作人员定期进行鸟类数量、生存状态的监测，基于这些监测数据进行统计分析形成的结果可以帮助保护区工作人员全方位、直观地掌握各种鸟类在整个青海湖不同地区、不同时间段内的分布、迁徙情况，有利于动物学家研究鸟类的活动规律，有利于鸟类保护工作的开展。本文结合保护区工作人员的具体需求，研究按地点、时间、物种等不同方式对青海湖的鸟类监测数据进行统计分析的方法，并介绍了该统计系统的功能、架构及实现细节。

关键词：青海湖；鸟类监测数据；统计分析；Amcharts

Design and Implementation of Qinghai Lake Wild Birds' Data Statistics Analysis System

Wan Sunhua, Huang Zhiyi, Zhou Yuanchun, Yan Baoping
Computer Network and Information Center, Chinese Academy of Sciences, Beijing 100190, China

Abstract: There are many kinds of wild birds live in Qinghai lake National Nature Reserve. Staffs of the nature reserve regularly count up the number of the birds, observe their activities. The processing results based on the analysis of the monitoring data should be useful the for the staffs to find out roundly and intuitively the situation of all kinds of birds which are distributed in the different areas of the Qinghai Lake at the different period. Also it is useful for the zoologists to study the law of the birds' distribution and migration, useful for the protection of the Qinghai Lake cross the birds. Based on the requirement of the reserve staff, this paper discusses several methods to analyze the data of birds in Qinghai Lake area, and introduces the method of statistical system function, structure and the implementation details.

Keywords: Qinghai Lake; Birds monitoring data; Statistic analysis; Amcharts

1. 引言

青海湖是中国最大的湖泊。青海湖位于青海省会西宁以西100千米，海拔3205米，是中国最大的咸水湖。有23条河流和小溪汇入青海湖，其中有很多是季节性河流。有5条主要的河流提供了青海湖总蓄水量的80%^[1]。

青海湖共有野生鸟类 1 8 9 种、1 5 万多只，近年来保持着稳中有增的态势。青海湖是国际重要保护湿地之一，是我国重要野生迁徙鸟的夏季和冬季栖息地。其中斑头雁、棕头鸥、鸬鹚、渔鸥是最常见的迁徙鸟类，数量都以万计。此外，赤嘴潜鸭、红脚鹬、环颈鸬等数量都在一千只以上。而燕鸥、白骨顶、凤头鸊鷉等数量较少，在青海湖地区分别有几百只。黑颈鹤分布在环湖湿地中，大约只有几十只，属于非常罕见的物种^[2]。近年来，有愈千只大天鹅在青海湖越冬。通过几年来的调查与研究，已积累了大量的青海湖地区主要鸟类分布的观测数据和图像，为保护和研究工作提供了有效的支持。

对青海湖的鸟类数量进行统计分析，是进行疾病传播监测、鸟类物种鉴别等研究的前序步骤，也是在获得珍贵的鸟类数据，并对其进行恰当的管理之后的有较高实用价值的应用。本系统对鸟类的数量，在时间、空间和物种等三个维度上进行细致的统计分析，得到鸟类物种数量的趋势图、比例图、柱状图和统计

表格等分析结果。系统采用MVC（Model模型-View视图-Control控制器）架构，对不同的分析条件进行建模，使得系统结构清晰，代码的冗余程度降到最小^[3, 4]。

2. 青海湖基础数据库

本文所讲述的鸟类数据统计分析系统建立在青海湖基础数据库之上，该数据库包含了字典数据、调查数据、鸟类GPS数据、病毒监测数据等。字典数据主要是包括青海湖调查数据、鸟类物种信息、植物、动物物种信息、鸟类跟踪数据及病毒等相关数据；调查数据是保护区工作人员定期巡湖获取鸟类数量、生存状况、植被生长状况等信息；鸟类GPS数据是绑在鸟身上的GPS仪器发回的GPS数据，记录了鸟类的飞行轨迹；病毒监测数据是由武汉病毒所等单位的科研人员定期采样监测的数据。本文中使用的数据主要是基础数据库中的字典数据和调查数据。

各类数据汇集后统一管理存储，并在全部数据内容的基础上根据学科研究主题进行相关数据资源的集成，使得用户能更有效地利用已有的数据，进行数据

分析，挖掘潜在的科学知识和信息。其中鸟类物种信息表的结构如表1所示。

3. 系统架构及功能设计

3.1 系统架构

通过分析青海湖保护区工作人员记录下来的监测数据，我们把统计对象分成：环湖调查数据、定点调查数据、黑颈鹤、普氏原羚4类数据。环湖调查数据是保护区工作者定期到鸟类的栖息地做的数量调查；定点调查数据是在蛋岛、鸬鹚岛每天定时记录的数据；黑颈鹤、普氏原羚由于其珍稀且保护级别高的原因，独立做监测记录及统计分析。

为了不同的可视化需求，本系统设计并实现了曲线图、柱状图、饼图三类统计图形。

曲线图：显示各采样地各种鸟类的在一段时间范围内的数量变化趋势图；显示某些采样地某些鸟类的数量总和（地点累加，鸟类累加，地点和鸟类都累加）。

柱状图：显示某段时间范围内各采样地鸟类数量情况；显示某段时间范围内，某地点各种鸟类数量情况。

饼图：显示各种鸟类的比

表1 青海湖野生鸟类数据表结构

名称	fields	类型与大小	概要字段
鸟类代码	BIRDID	Int(10)	
名称	C_SCIENTIFICNAME	Varchar100	√
英文名称	E_SCIENTIFICNAME	varchar40	√
拉丁学名	ACCEPTEDNAME	varchar40	√
.....	

例情况；显示各采样地鸟类的比例情况。

系统采用MVC模式设计，使用了java开源框架spring、struts来组织代码结构，并使用amcharts框架和jQuery前端开源框架来进行数据的最终展示。数据统计分析系统的系统流程图如图1所示。

3.2 功能设计

系统可以实现在时间、地点、物种三个不同维度上对数据进行统计和分析。在时间维度上，用户可以设定年、月、日三种不同的时间粒度，使用户能得到在不同精度和时间跨度上的数据统计结果。在进行地点的选择上，系统采用了GOOGLE MAP对用户选择的地点在地图上进行实时显示，以此来增强用户体验。而在鸟类物种选择时，系统采用AJAX架构，从系统后台数据库中实时读取鸟类信息，并将鸟类的种名、所属的目、科，以及图片显示在网页上，使用户能对所分析的物种有全面了解。系统的查询条件输入界面如图2所示。

系统按照功能模块分可以分为曲线图、柱状图和饼图三种显示方式。系统曲线图用于显示各采样地各种鸟类的数量，以及显示某些地点某些鸟类的数量总和（地点累加，鸟类累加，地点和鸟类都累加）等。柱状图用于显示各地鸟类数量情况以及各种鸟类数量情况。而饼图则用于显示各种鸟类的比例情况以及各地鸟

类的比例情况。

系统按照查询模块来划分，可分为黑颈鹤数据、环湖调查数

据、定点调查数据等功能模块。

系统的查询结果示例展示如图3~图5所示。除了展现统计图以

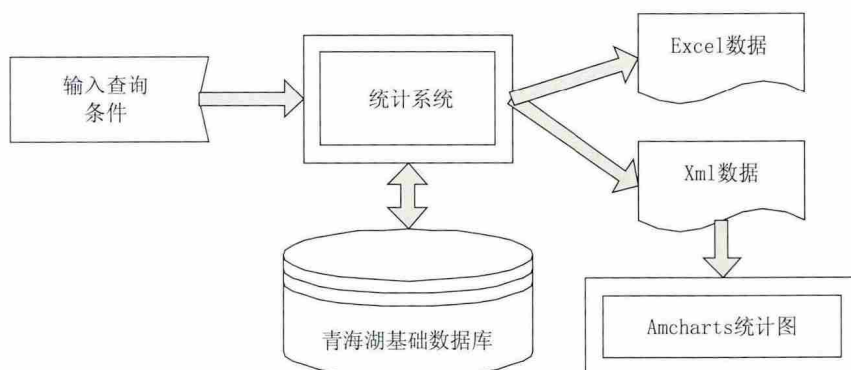


图1 系统流程图

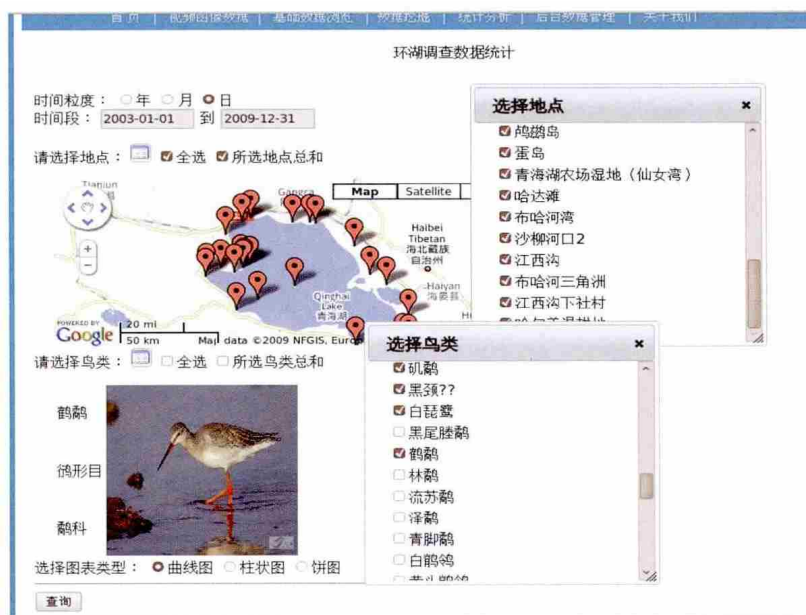


图2 查询界面概览

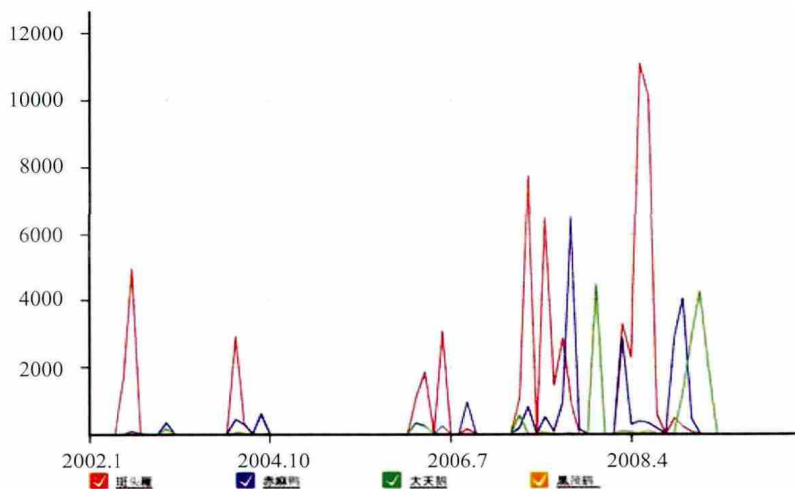


图3 环湖调查曲线图（按鸟类比较）

外, 系统还提供查询统计数据下载、统计图表导出等图形文件的功能。

图3显示了在2003年1月至2008年12月期间, 以月为时间粒度单位, 斑头雁、赤麻鸭、大天鹅、黑颈鹤四种鸟类的数量变化趋势图(2005年没有调查, 记录为空白的)。从图中可以看出青海湖是大天鹅冬季栖息地之一, 是斑头雁的春夏季栖息地之一, 2008年斑头雁的数量达到了1万只以上。

图4显示的是洱海、沙柳河口、铁卜加河口、小湖泊四个地点在同一时间段内, 鸟类总量的对比情况, 从图中可看出小湖泊的鸟类数量最少。

图5-1和图5-2分别是2008年5-6月和11-12月在青海湖巡湖调查的所有调查点记录下的斑头雁、赤麻鸭、大天鹅、黑颈鹤、棕头鸥、鱼鸥6种鸟类的数量统计情况。从中可看不同鸟类在不同季节在青海湖的统计数量。

4. 统计系统的实现

系统采用了MVC架构来组织代码结构。为实现MVC架构, 系统使用到的框架有spring、hibernate、struts、amcharts等。系统使用到的java技术包括特定格式的XML文件的生成; EXCEL等格式文件的读取、生成、保存和下载; 数据库编程; Ajax技术等^[5, 6]。

用户查询界面涉及到查询

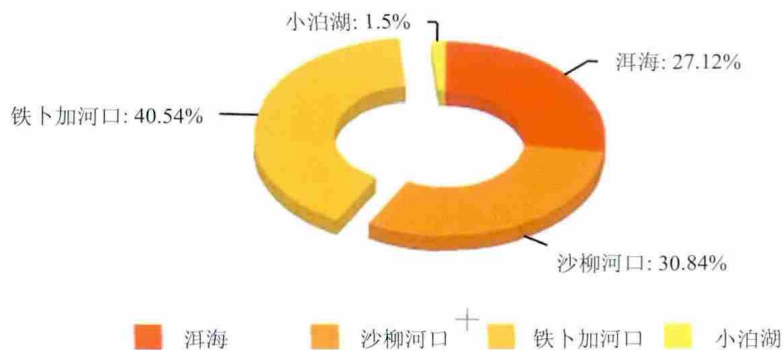


图4 环湖调查饼状图 (各地数量比)

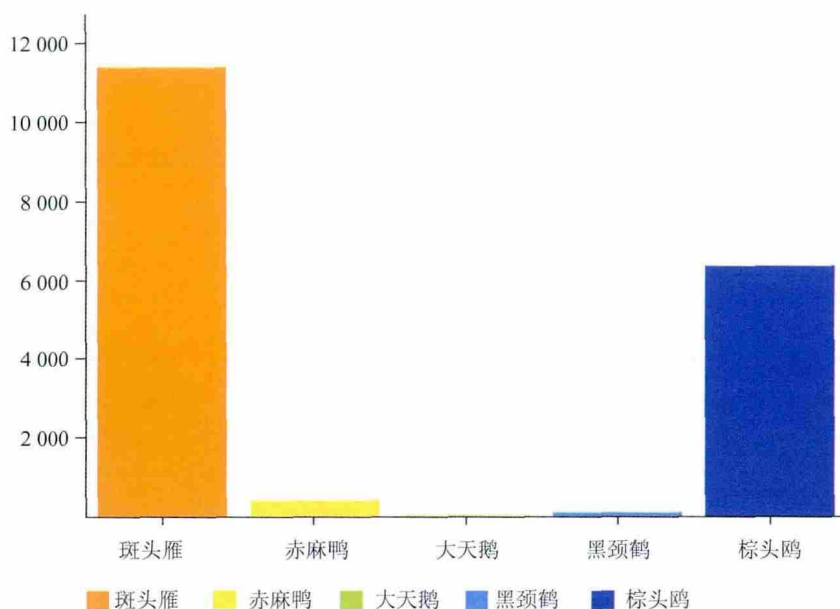


图5-1 环湖调查柱状图2008年5-6月 (按鸟类比较)

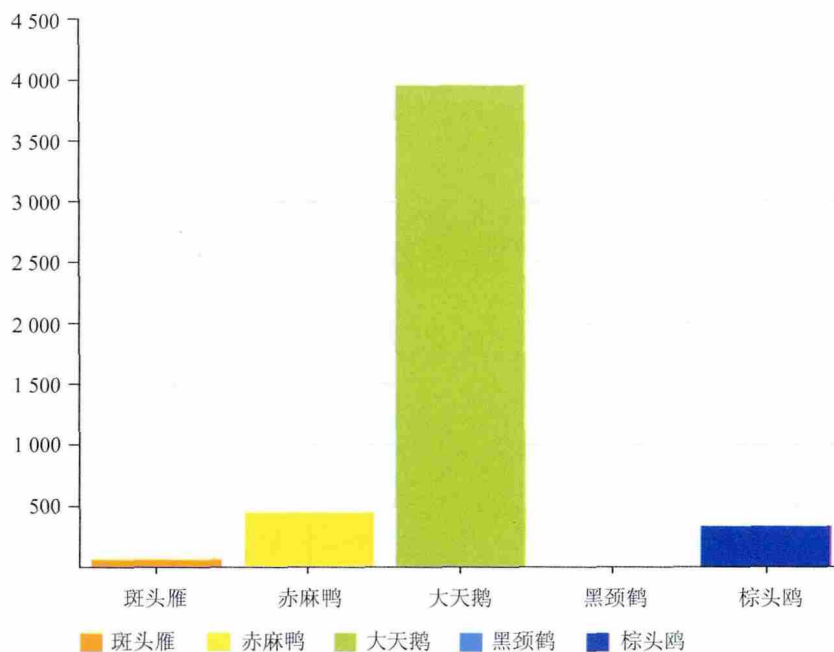


图5-2 环湖调查柱状图2008年11-12月 (按鸟类比较)

条件输入方式（直接输入、复选框、下拉菜单等）；查询时页面跳转等问题。系统在表现层采用Ajax技术，从数据库提取现有数据供用户选择，使用户在查询时页面不发生跳转，提高了用户的体验。

系统在实现过程中，根据以往项目使用Hibernate的经验，考虑在本项目中采用Hibernat, 重点利用了Hibernate以下优点：

1. 将实体映射到数据库中，不需以手动的方式进行持久化存储了。

2. Hibernate实现了数据的存储、加载、缓存等方面的工作。

3. Hibernate没有侵入性（轻量级框架），编写的类不用继承Hibernate的任何类。

在本系统中，Hibernate主要用于提取单个的对象，对这个对象进行修改、存储的情况。

不适合Hibernate的情况包括：需要对数据进行批量的读取修改和存储；OR映射过程太复杂；要使用到sql语句的特定功能，其原因是由于在使用Hibernate时，生成的sql语句不可控，这将导致过于具体化的需求无法在Hibernate框架中得以体现^[7, 8]。

本系统在更新过程中进行过一次代码结构的重组，以使得码的冗余度最小，并且利于对代码进行扩充和适应相应的需求变更。系统结构如图6所示。

在使用Spring框架对系统进行代码重构之后，系统的架构如

图7所示。在重构之后，前端使用Ajax技术，避免了页面跳转，使得用户查询的等待时间减少。而后端代码组织时，使用Spring技术，使得系统在四个方面的代码可以复用，这样可以在实现复杂功能时，继承之前的简单功能的实现类，增强了代码的可扩展性。

在系统的实现过程中，解决了一些遇到的技术上的问题。代码重复的问题是初期时遇到的最大的问题。系统由四个功能相似的部分组成，但是这些功能所基于的数据表结构不同，且查询条件也不尽相同，使得编码时重复编写相似代

码并导致最后代码产生大量冗余。改进后的系统采用的解决方案是使用java面向对象的继承机制，将公共部分提取出来^[9]。

系统涉及到用户的不同查询方式的组合。由于系统的查询输入条件较多，导致有很多不同的查询条件组合方式。系统采用的解决方式是，提出了一些实用性较强的查询结果，以此来组合查询条件。

最后，在系统需求变更控制上，开发过程中采用的控制方法是，将功能分层分块，以降低需求变更带来的影响。

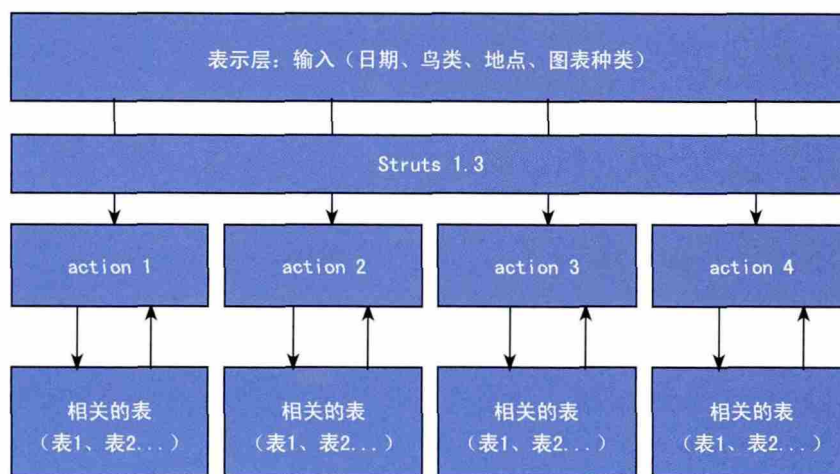


图6 数据统计分析系统原系统结构

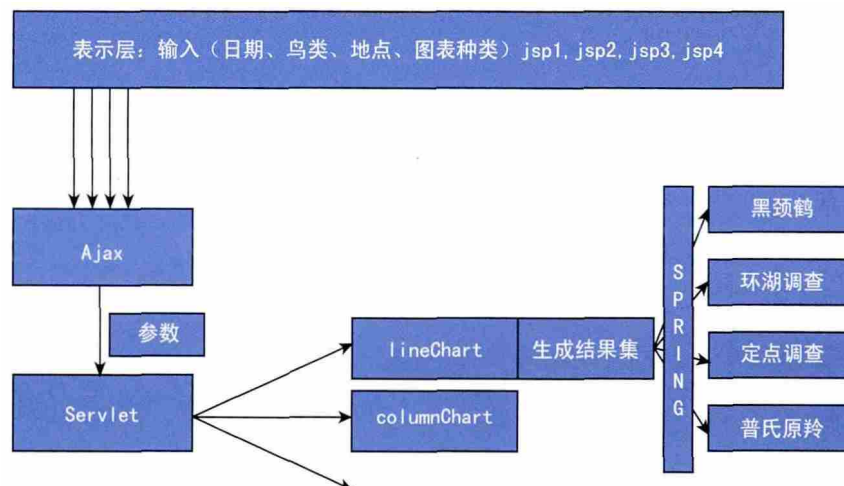


图7 数据统计分析系统改进后系统结构

►5. 结束语

青海湖鸟类数据统计分析系统为青海湖保护区的工作人员掌握鸟类的数量、迁徙、分布、生存状况提供了重要帮助，同时也

为鸟类科研工作者提供了宝贵的一手资料。当然这也是建立在保护区工作者不辞辛苦的进行各种监测调查活动，并将记录下来的数据入库的基础之上的。在这一系列科学数据的产生、收集、处

理、分析、挖掘过程中，我们还可以做更多的事情帮助科研人员进行科学研究，使得保护区工作者、相关科研人员更快捷的利用这些科学数据，进行科学的保护和科学的研究与发现。



参考文献:



- [1] Li, X.Z., et al., Distribution of Recent ostracod species in the Lake Qinghai area in northwestern China and its ecological significance. *Ecological Indicators*, 2010, 10(4): p. 880-890.
- [2] Lvov, D.K., et al., Evolution of Highly Pathogenic Avian Influenza H5N1 Virus in Natural Ecosystems of Northern Eurasia (2005-08). *Avian Diseases*, 2010, 54(1): p. 483-495.
- [3] Zhou, J.Y., et al., Characterization of the H5N1 Highly Pathogenic Avian Influenza Virus Derived from Wild Pikas in China. *Journal of Virology*, 2009, 83(17): p. 8957-8964.
- [4] Minta, Z., et al., Highly pathogenic avian influenza H5N1 in wild birds in Poland - analysis of first cases. *Medycyna Weterynaryjna*, 2007, 63(11): p. 1349-1352.
- [5] Ni, F.-y., et al., The system framework of university's opening experiment management based on MVC model. *Hunan Shifan Daxue Ziran Kexue Xuebao*, 2008, 31(2): p. 35-39.
- [6] Dai, W. and S.J. Xue, The Research of Students Management Information System Based on J2EE. *Dcables 2008 Proceedings, Vols I and II*, 2008: p. 837-842.
- [7] Tili, M., et al., A web-based architecture for e-gov application development. *Ice-B 2008: Proceedings of the International Conference on E-Business*, 2008: p. 476-479.
- [8] Chen, E.X. and M.H. Liu, Research and Design on Library Management System Based on Struts and Hibernate Framework. *2009 Wase International Conference on Information Engineering, Icie 2009, Vol II*, 2009: p. 310-313.
- [9] Agostinho, S., P. Guerreiro, and H. Taborda, An aspect for Design by Contract in Java. *Msvveis 2008: Modelling, Simulation, Verification and Validation of Enterprise Information Systems*, 2008: p. 119-128.

收稿时间：2010年10月8日

作者信息



万孙华

中国科学院计算机网络信息中心，硕士研究生。主要研究方向为数据挖掘。



黄志一

中国科学院计算机网络信息中心，助理工程师。主要研究方向为移动数据采集、数据密集型计算。



周园春

中国科学院计算机网络信息中心，副研究员，博士。主要研究方向为数据挖掘、数据密集型计算、科学工作流。



阎保平

中国科学院计算机网络信息中心，研究员，博士生导师。主要研究方向为e-Science应用示范，大规模数据集成和处理。