

科技管理数据挖掘和基于 WebGIS 的展示

严会超¹, 吴小芳², 王璐², 陈飞香²

(1 华南农业大学 科技管理与规划研究所, 广东 广州 510642; 2 华南农业大学 信息学院, 广东 广州 510642)

摘要:在分析传统的科技管理工作不足的基础上,设计了一体化的科技管理数据库,通过引入 workflow 技术、数据挖掘技术、GIS 技术,着重研究了项目申报审批与工作的流程化定制、GIS 图形可视化表达和数据挖掘与统计分析等关键技术,构建了基于 WebGIS 的科技管理信息系统平台。

关键词:GIS; 数据挖掘; 统计分析

中图分类号:TP311

文献标识码:A

文章编号:1001-411X(2009)02-0101-04

Data Mining and Demo of Science and Technology Management Information System Based on WebGIS

YAN Hui-chao¹, WU Xiao-fang², WANG Lu², CHEN Fei-xiang²

(1 Research Institute of Science & Technology Management and Planning, South China Agricultural University, Guangzhou 510642, China; 2 College of Information, South China Agricultural University, Guangzhou 510642, China)

Abstract: With the disadvantages of traditional science and technology management system analyzed, an integrative science and technology management database was designed. Data mining, workflow and geographic information systems (GIS) technology were employed to study key problems in detail, such as workflow of project, map visualization, data mining and statistical analysis. Then, the science and technology management information system based on WebGIS was developed and constructed.

Key words: GIS; data mining; statistical analysis

科技管理是科技活动中一个非常重要的环节,如何高效、准确地了解科技活动动态,掌握科技发展的概况,及时了解科技研究进展,一直是科技管理者十分关注的问题^[1]。传统的科技管理工作很大程度上依赖于人工进行,对科技信息的采集和整理比较困难,从而加重了科技管理人员的负担,减少了与科技人员之间有效的沟通,科技资源得不到合理的配置和利用。在信息化技术日益普及的今天,通过成熟的计算机技术,实现科技管理信息化,加强科技工作管理,整合科技工作流程,为决策层提供真实有效的参考依据,从而提升科技管理水平,是当今科技管理的大势所趋^[2-4]。本文在分析传统的科技管理工作不足的基础上,设计了一体化的科技管理数据库,通过引入 workflow 技术、数据挖掘技术、GIS 技术,着重研究了项目申报审批与工作的流程化定制、GIS 图形可视化表达和数据挖掘与统计分析等关键技术,构建了基于 WebGIS 的科技管理信息系统平台。并以国家

自然科学基金委员会的项目数据为例,研制原型系统,从时间轴、空间轴等多角度进行空间数据挖掘分析,挖掘蕴藏在项目数据中的各种知识和规律,从而为科技管理提供有效的研究平台和决策依据。

1 数据库设计

本文通过对科技管理的业务流程和系统需求进行深入分析,将科技管理数据库划分为数据编码库、科技信息专题数据库、基础地理数据库、业务流程库、业务专题图库五大数据库(图1)。

(1)数据编码库:是指采用统一的编码方式表达科技管理的部分专题数据的数据库,包括:学科编码、职称编码、民族编码、期刊编码、项目类型编码等。数据编码库的构建使得数据更加统一化、标准化,便于数据的修改、维护,减少数据冗余,提高程序编码的重复可用性,且为科技管理的数据挖掘及决策分析提供有效的支撑。如项目验收时,根据项目中

发表论文所属的期刊编码,可快速获取期刊的名称、否达到验收标准要求. 期刊的级别等信息,以此判断论文的质量(级别)是

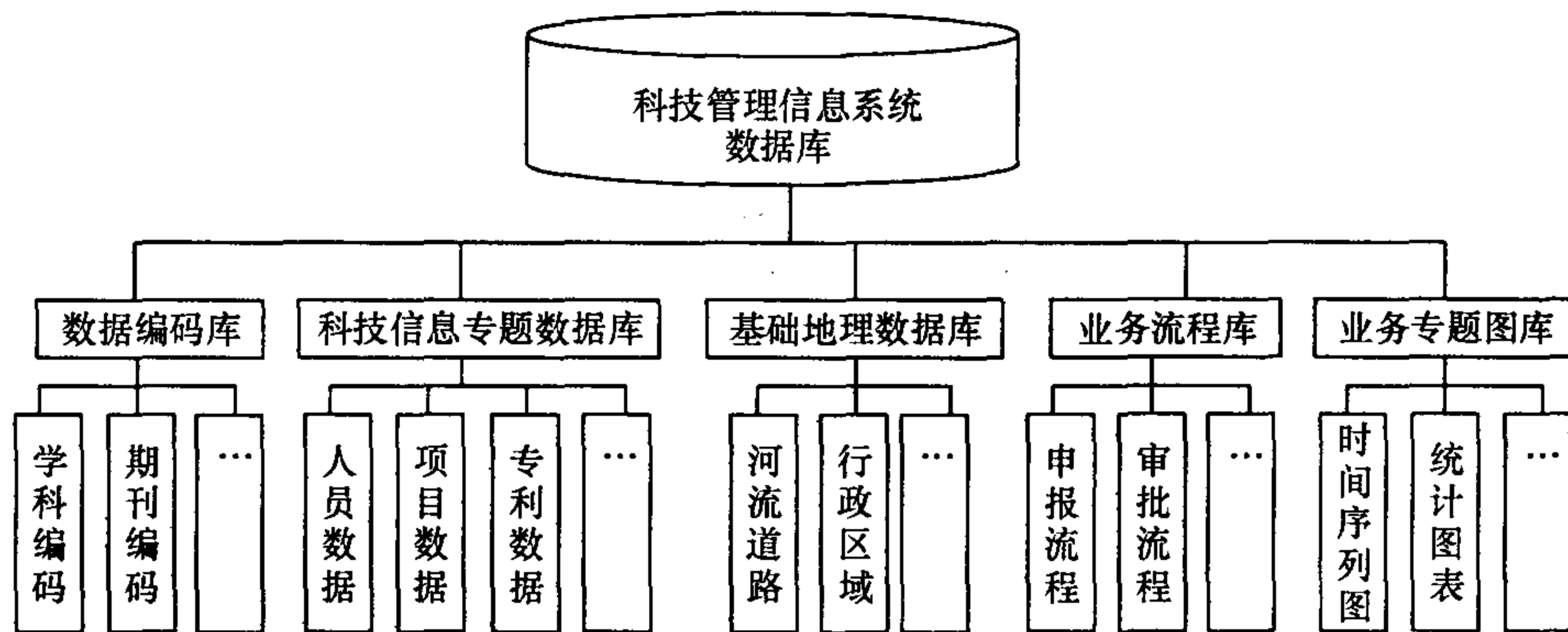


图1 科技管理信息系统数据库结构图

Fig.1 Database structure of science and technology management information system

(2)科技信息专题数据库:主要包括科技信息管理系统的日常业务管理所涉及的专题数据,如人员数据、项目数据、专利数据、成果数据、论文数据,著作数据等.每一类数据在数据库采用表结构的方式进行管理与维护.如人员数据表,其所涉及的字段信息为:姓名、性别、出生年月、所在单位、学历、职称、职务、从事专业、熟悉专业、联系电话、备注等.数据库中各表之间通过名称或ID相互关联^[5].

(3)基础地理数据库:包括常规的行政区划、公路、铁路、河流、境界等空间地理要素,其与科技信息专题数据进行关联,用于科技专题数据的空间定位、查询及可视化显示.

(4)业务流程库:是指科技信息管理系统中为方便各地区科研申报所设定的不同的申报流程库.建立该库后,只需要应用流程定制工具对该库进行灵活地构建与维护,即可满足不同地区的科研项目申报与审批的需求.

(5)业务专题图库:是管理各类特定专题图的

数据库,通过对该库进行维护,即可方便地实现新增专题图,或对已有专题图的样式、等级、符号等进行修改与维护.

在这五大库中,数据编码库是基础,科技信息专题数据库是核心,基础地理信息库是项目专题数据空间定位的数据来源,业务流程库是项目申报与立项业务的重要支撑,业务专题图库则是各项统计分析 & 可视化表达的依据.

2 系统设计与实现

2.1 系统整体架构

引入 WebGIS 技术,构建基于 WebGIS 的科技管理信息系统.采用 B/S 体系结构,以 SQL Server 为数据库管理平台,应用 .Net 开发平台,通过 ADO 技术和 GIS 的图形可视化、统计分析 & 关联规则数据挖掘技术,实现了基于 WebGIS 的科技管理信息系统.系统整体架构见图 2.

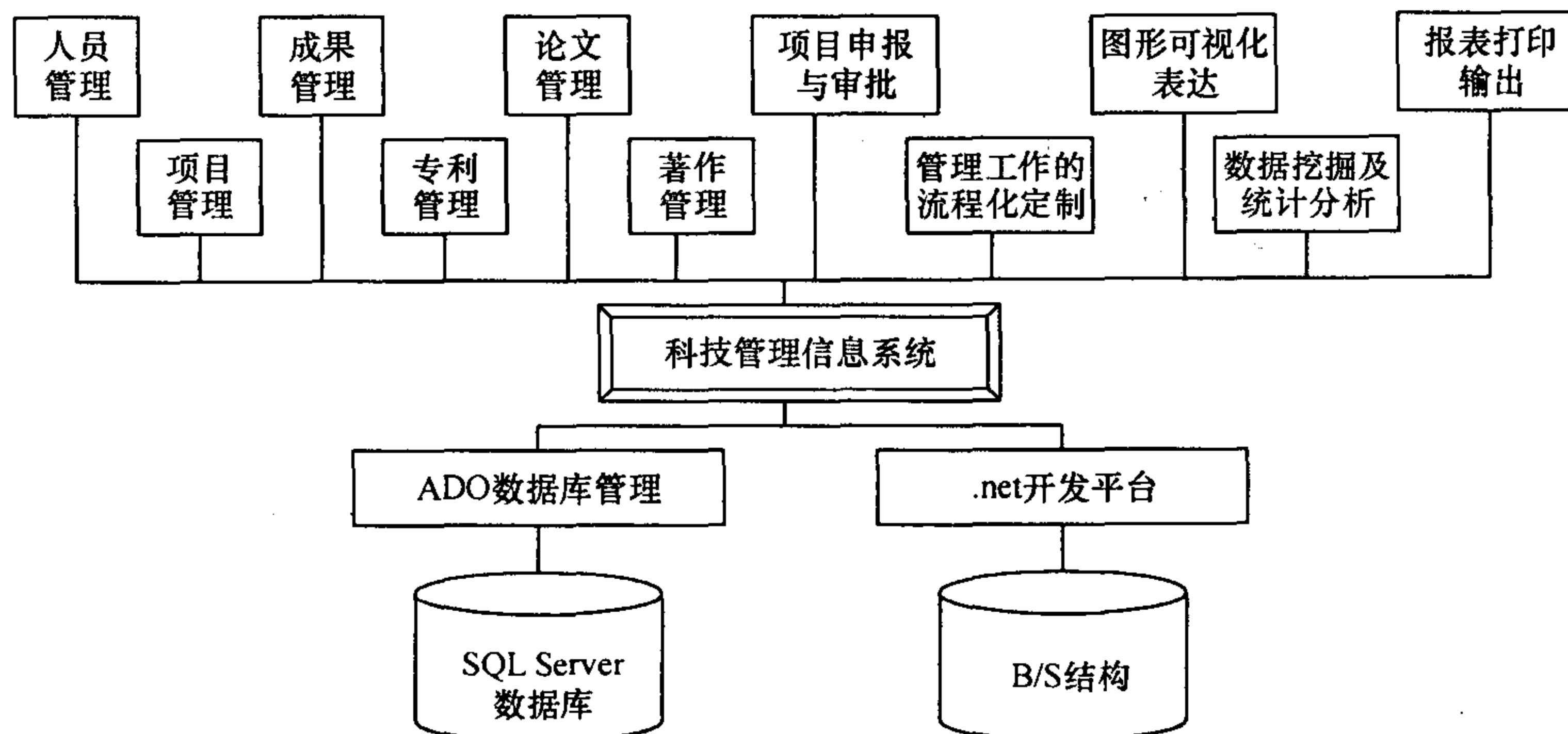


图2 系统整体架构图

Fig.2 General structural chart of system

2.2 项目申报审批与工作的流程化定制

项目申报与审批一般包括项目的开始申报、所在部门审查、科技主管部门审查、评审专家评审、主管部门审批、项目立项等多个环节,每一个环节的用户需在相应的申报表中填写相关的审批意见,为此,系统将该审批过程抽象为相应的工作流程图,将申报表抽象为表单,建立相应的流程库(包括表单库),并采用 XML 技术实现了项目申报审批的流程化。

在系统具体实现时,考虑到该系统面向多层面、多用户,不同用户的权限不一样的特点,针对每一环节(流程图中的节点),均提供了对表单中某些内容的编辑权限配置;同时,考虑到特定科技领域的申报流程有所差异,为适应业务的变化,系统运用业务流程库,采用了流程自定制技术与 XML 技术,提供了可视化的流程界面设计方式,可动态配置工作化流程,即可对流程图中的结点和链接进行增加、删除和修改,实现了工作流程的动态调整,满足了不同类型项目的业务申报。

2.3 图形可视化表达

图形可视化表达实现的基本思路是利用 GDI+ 技术,依据一定的地图符号,分别绘制地图的各个图层中所有要素,而后采用绘制线、矩形、文字等方法,绘制相应的专题图表要素,并在服务器端生成 Image 图像,最后将其发送到客户端,进行图形的显示。

图3是国家自然科学基金项目中,各省市项目数据的柱状专题图形可视化表达。实现过程是:根据客户端用户的选择,动态提取记录集,然后利用记录集的 ProvinceName 字段与省市图层的 Name 字段进行实时关联,然后依据各省市要素的中心点作为定位点,根据记录集中的各专题数据值绘制的柱状图,生成 Image 图像,将其叠加在整幅地图图形上,形成专题地图的可视化表达。

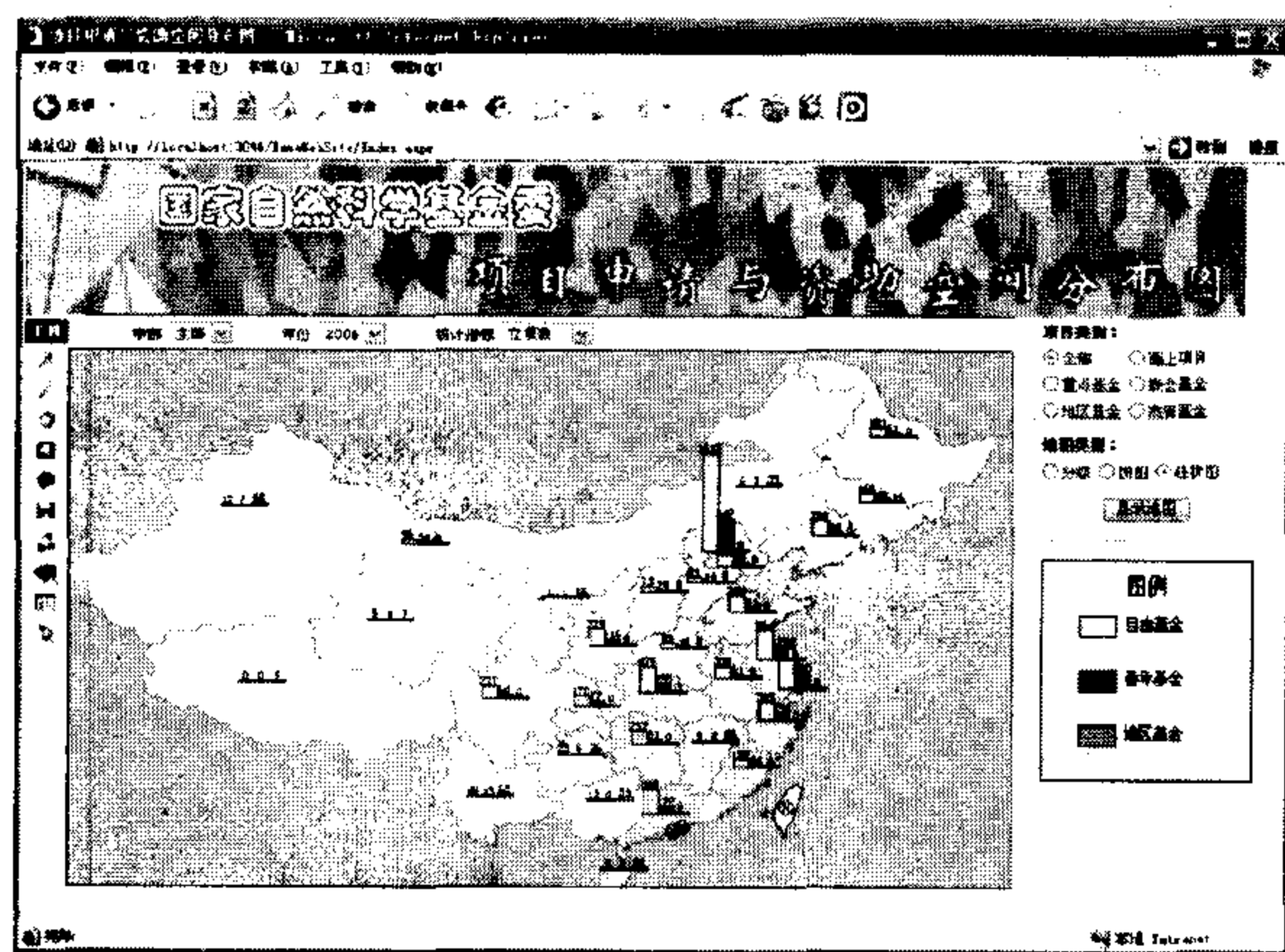


图3 各省市项目数据的专题图形可视化表达

Fig. 3 Visualization of thematic map on the number of projects of all provinces

3 系统应用分析

3.1 从时间轴分析

主要是从时间上进行分析,展示不同年度、不同区域、不同类型的科研立项数变化,从中发现国家在各学科科研项目上随时间的变化发展规律。

以2005—2007年各省市的面上基金、青年基金、地区基金项目数据为基础数据,结合各省市的地理位置及空间分布,分别统计出我国东部、西部、中部3类基金的立项总数,并采用统计图表的方式,展现东部、中部、西部的立项数随时间的发展规律。从图4中可知,从2005—2007年,东部、中部、西部的立项数目呈逐年递增的趋势变化,其中,西部与中部地区的立项数目递增速度相当,东部地区的立项数目递增速度较快。此外,东部地区的立项数最多,这也反映了我国东部较西部地区的科研力量强的现实。

在此基础上,以2007年数据为例,详细展示东部、中部、西部地区不同类型基金的立项数,以此分析不同区域的各项类型项目的立项优势。从图5中可知,东部和中部地区主要以面上基金、青年基金为主,地区基金项目数较少,而在西部地区,其地区基金项目数较高,占据一定的比例,这也体现了国家对西部发展的地区基金政策的扶持力度。

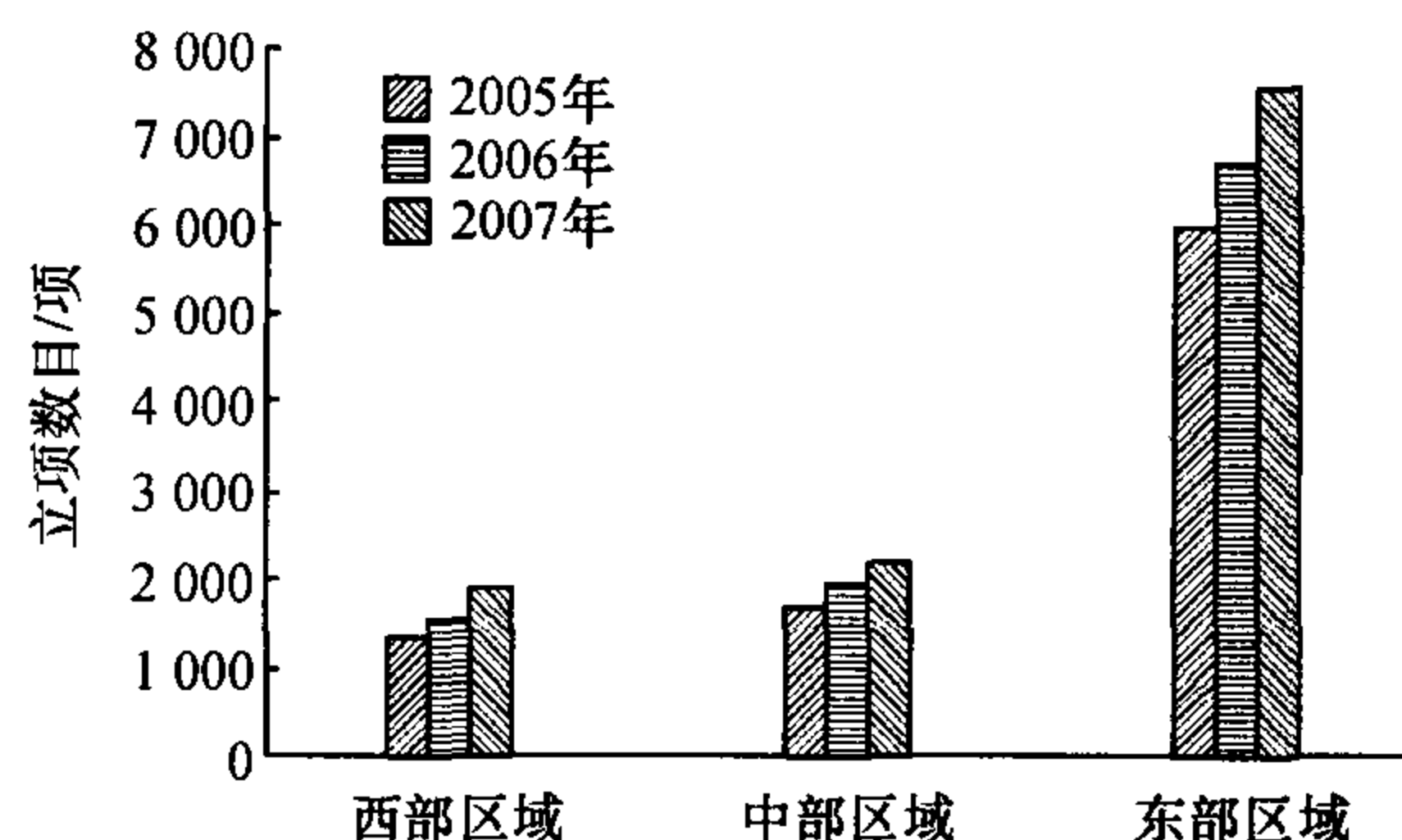


图4 立项数目变化图

Fig. 4 Chart of number of projects

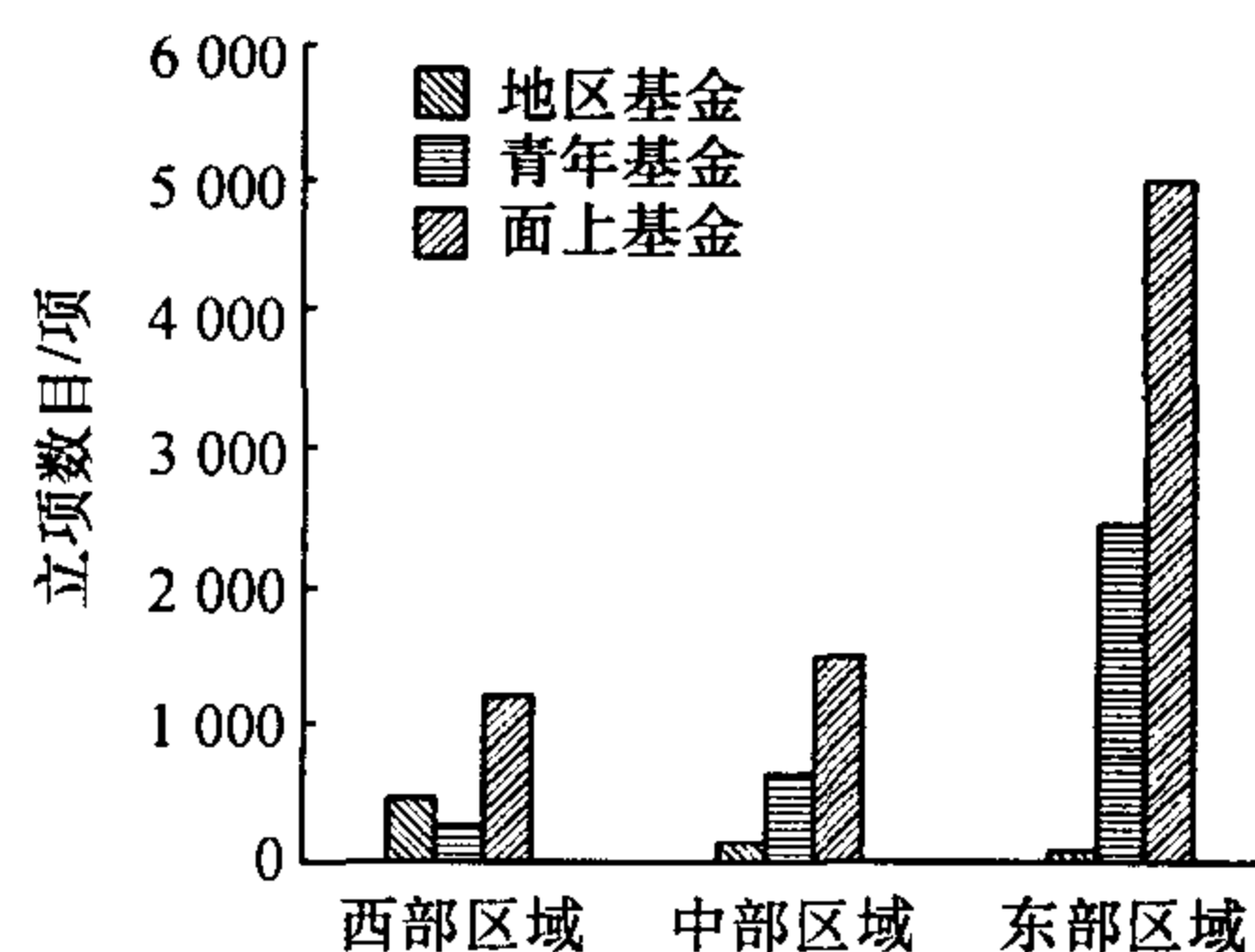


图5 2007年各类型项目分布图

Fig. 5 Chart of types of projects in 2007

3.2 从空间轴分析

图6是采用分级设色的专题地图表达方法,展示各省市的面上项目立项数分布情况.以各省市为区域单位,将项目立项数目分为多个等级,并对每一等级设置不同的颜色,以展示各省市的面上项目立项数量情况.从图6中可以发现,北京市作为我国高校和科研机构集聚的地区,其立项数目为全国各省市最多,其次是江浙一带的沿海城市,以及中部地区高校较集中的湖北省和陕西省,再次是西部地区.通过图形的直观对比分析,可以得出,以北京市为首的东部地区,科技项目力量最强,中部地区次之,西部地区较弱.因此,可考虑西部地区与中、东部地区结合,联合申报项目,以带动西部科技发展,促进西部的科研能力和科研团队的形成,推动西部科学研究的发展.

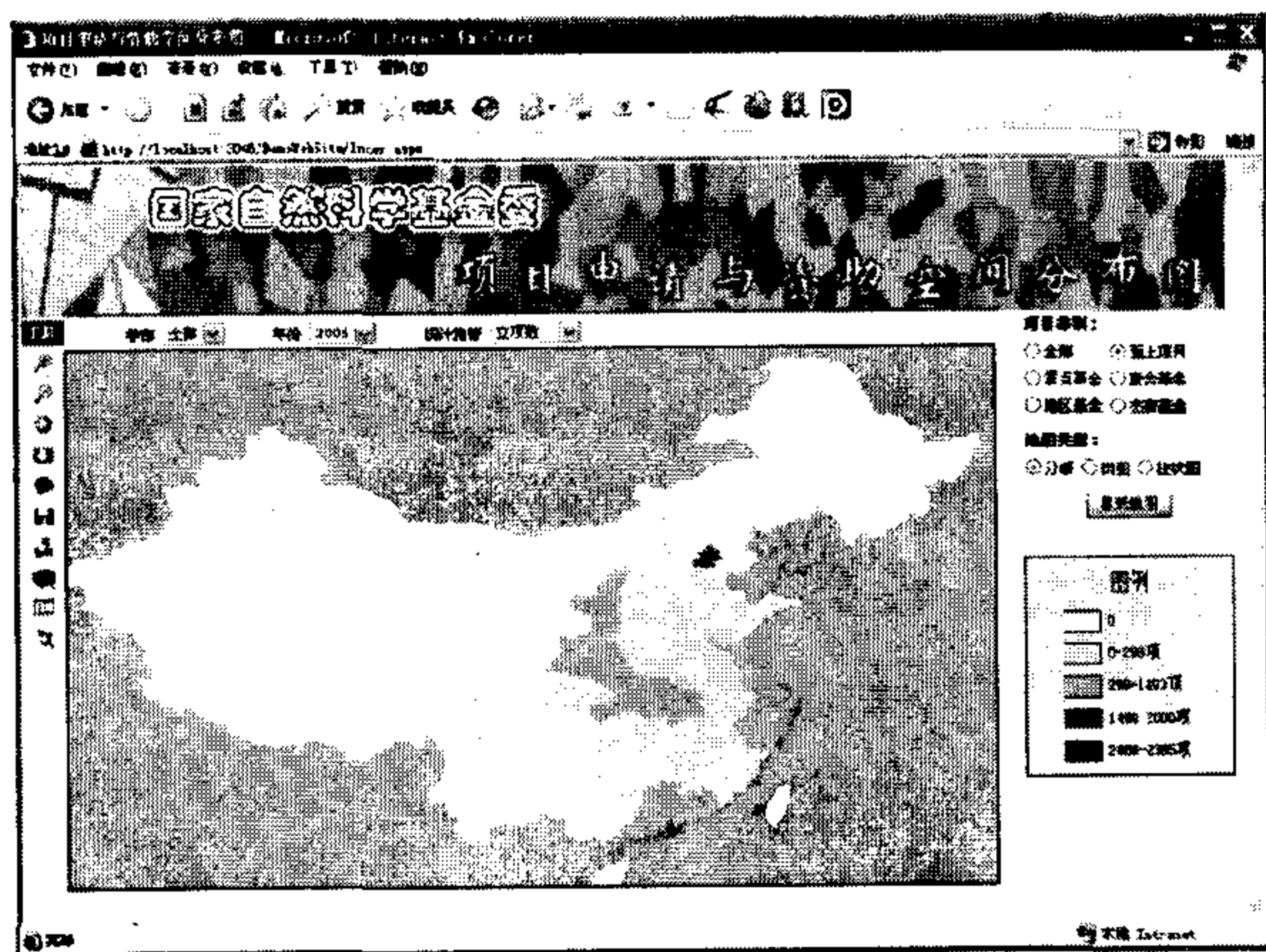


图6 项目空间区域分布图
Fig.6 Spatial distribution map of projects

3.3 从项目专业类型轴分析

图7采用柱状图专题表达方法,展示了面上项目中各不同学部的立项项目数,从图7中可以看出,生命科学部作为国家基金委中项目比较多的学部,相对于其他学部,其在各省市的立项数目均是最多的,占据着绝对优势,其他学部如:数理科学部、化学科学部、地球科学部、工程与材料科学部、信息科学部次之,而管理科学部的立项数相对最少.因此,在保持生命科学部中优势学科的基础上,需加强对其他学部及学科的支持力度.

此外,系统还实现了从项目负责人年龄及学历轴等多种角度进行分析,展示不同年龄及学历的项

目负责人科研立项数,从中发现科研项目立项与负责人的年龄、学历、职称之间的相关规律,从而为科技人员的项目申报提供辅助支持.

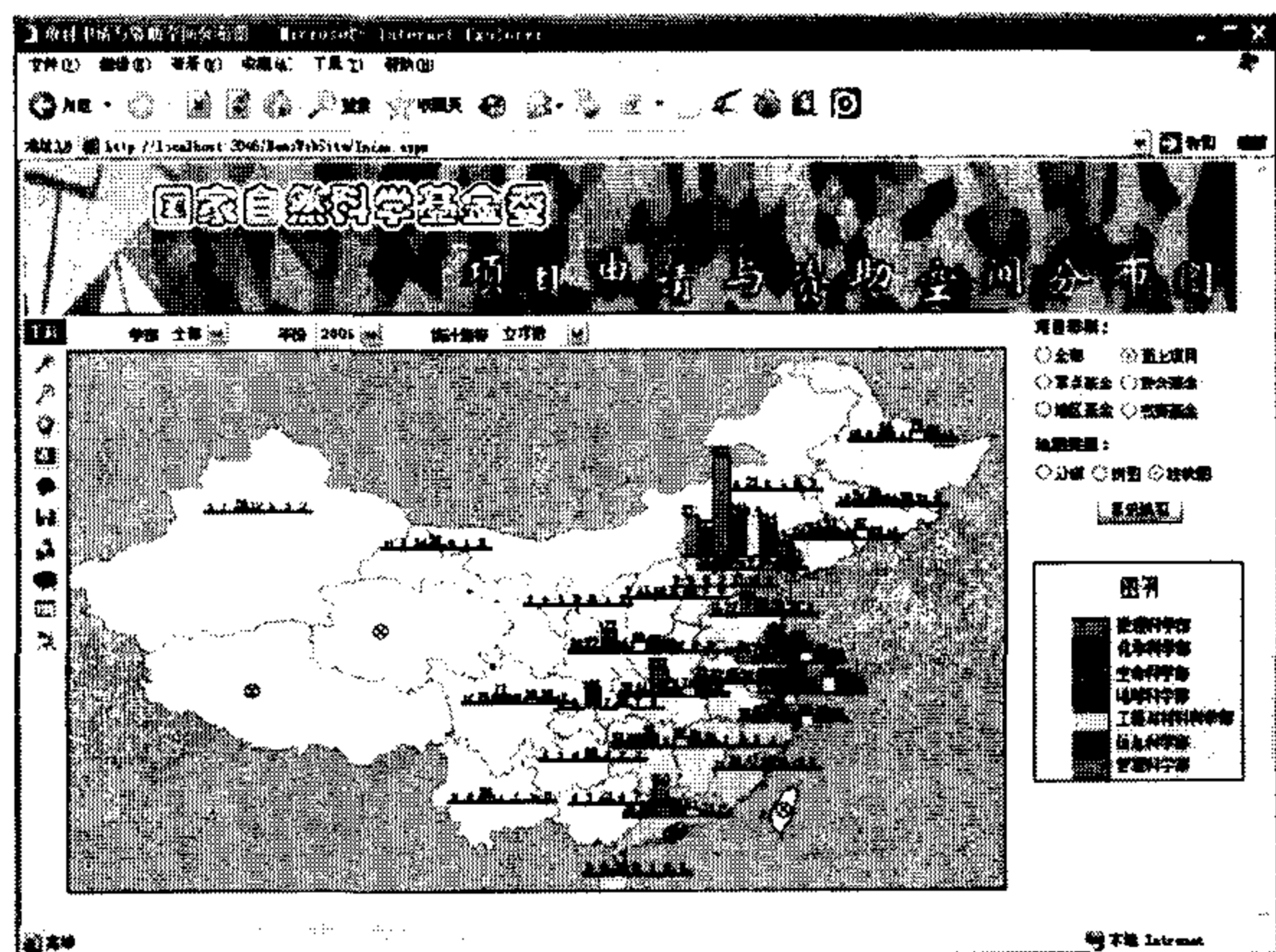


图7 各专业项目数的专题图
Fig.7 Thematic map of projects by specialties

4 小结

基于 WebGIS 的科技管理信息系统的构建能够最大限度地提高科技管理的工作效率,建立协调、规范的管理体系,实现对科研情况有效的管理以及对科研活动的全程跟踪,并可考虑与教育部等部门科研管理系统的数据对接,既实现对科研活动的高效率管理和控制,又满足决策者对科研活动的宏观管理与决策的需要,有效地促进科研信息管理的远程化、现代化和标准化.

参考文献:

- [1] 杨莉,徐俊利,贾广雷.基于 Web 的科技信息管理系统分析与设计[J]. 山东建筑大学学报, 2007, 22(5): 464-466.
- [2] 倪江林.科技计划项目管理信息系统研制与应用[J]. 企业技术开发, 2008, 27(2):16-18.
- [3] 乔祥源.关联规则挖掘算法研究及在科技管理信息系统的应用[D]. 南京: 南京理工大学计算机科学与技术学院, 2007.
- [4] 钱怡.高校科研管理信息系统的研究与开发[J]. 华北工学院学报, 2004, 25(5):353-355.
- [5] 彭晓晖,王丛刚,瞿裕忠.一个基于 Web 的高校科技信息管理系统[J]. 计算机应用软件, 2003, 20(6):31-33.

【责任编辑 李晓卉】