

1991—2015

GUOJIA SHEHUI KEXUE JIJIN JIQI
XUESHU CHENGGUO TONGJI BAOGAO

国家社会科学基金 及其学术成果统计报告

王东波 白云 朱丹浩 姜霖 著

国家社会科学基金及其学术成果 统计报告:1991—2015

王东波 白云 朱丹浩 姜霖 著



國家圖書館出版社
National Library of China Publishing House

图书在版编目(CIP)数据

国家社会科学基金及其学术成果统计报告:1991~2015/王东波等著.--北京:
国家图书馆出版社,2015.12

ISBN 978-7-5013-5362-0

I. ①国… II. ①王… III. ①社会科学—研究成果—
研究报告—中国—1991~2015 IV. ①C12

中国版本图书馆 CIP 数据核字(2015)第 311393 号

书 名 国家社会科学基金及其学术成果统计报告:1991—2015
著 者 王东波 白 云 朱丹浩 姜 霖 著
责任编辑 王 雷 王炳乾

出 版 国家图书馆出版社(100034 北京市西城区文津街7号)
(原书目文献出版社 北京图书馆出版社)

发 行 010-66114536 66126153 66151313 66175620
66121706(传真),66126156(门市部)

E-mail nlcpress@nlc.cn(邮购)

Website www.nlcpress.com →投稿中心

经 销 新华书店

印 装 北京科信印刷有限公司

版 次 2015年12月第1版 2015年12月第1次印刷

开 本 787毫米×1092毫米 1/16

印 张 23.75

字 数 500千字

书 号 ISBN 978-7-5013-5362-0

定 价 100.00元

本书出版获得以下研究项目的资助:

国家自然科学基金青年项目(71303120):基于 CSSCI 的句法级汉英平行语料库构建及知识挖掘研究

江苏省社科基金青年项目(13XWC017):大数据环境下汉英短语级平行语料标注及知识挖掘研究

2011 湖北省协同创新中心项目(JD20150101):面向应急推演平台的海量突发事件知识库与模型库构建研究

前言

自1991年国家社会科学基金(简称国家社科基金)设立以来,无论是在立项的数量上还是所产生的成果上均达到了一定的规模。如何对国家社科基金及其成果从历时的角度进行定量统计与分析,以便于更好地发挥国家社科基金在“推进我国学术理论创新、为协调推进‘四个全面’战略布局、促进经济社会持续健康发展提供有力理论支撑”历史进程中的重要作用,逐步被提上了议事日程,正如中央政治局委员、中央宣传部部长刘奇葆在2014年国家社科基金项目评审工作会议上所强调的,“要加强国家社科基金管理工作,创新选题规划、立项资助、成果评估和推广应用机制”,其中“成果评估”作为重要的一方面便位列其中。

在上述这一大的背景下,结合信息计量学的方法,基于国家社科基金项目数据库中的项目基本信息和中国知识基础设施工程(CNKI)中国家社科基金资助的学术论文,本专著主要从国家社科基金的基本信息、国家社科基金标题内容分析、国家社科基金的结题情况、国家社科基金的学术研究成果四个方面对马列·科社,党史·党建,哲学,理论经济,应用经济,政治学,社会学,法学,国际问题研究,中国历史,世界历史,考古学,民族学,宗教学,中国文学,外国文学,语言学,新闻学与传播学,人口学,统计学,图书馆、情报与文献学和体育学22个人文社会科学的国家社科基金及其研究成果进行了统计和分析,从整体上勾勒出了1991—2015年间的国家社科基金发展概貌。有关国家社科基金统计和分析的四个方面具体如下:

首先,国家社科基金基本信息方面。在获取到的国家社科基金基本信息的基础上,本专著主要从社科项目年度分布、项目类型分布、项目负责人和项目承担单位等几个方面进行统计和分析。针对项目负责人的相关信息,在统计的过程中又对项目负责人出现的频次、项目负责人的职称情况进行了细化的统计和分析。

其次,国家社科基金标题内容方面。基于内容分析的方法,结合自动分词的技术,本专著以汉字和词汇为基本单位,统计了22个学科国家社科基金项目标题的长度,并对标题的长度分布情况进行了分析。在对标题自动分词的基础上,本专著统计了标题中常用的汉语词汇,并依据词汇出现的频次,分析了不同学科项目标题中常用的词汇情况。

再次,国家社科基金的结题情况。根据国家社科基金数据库中的结项信息,本专著主要从国家社科基金的结项时间跨度、结项专著的出版机构统计和结项评定结果三个方面进行了统计和分析。在时间跨度的长度选择上,本专著对于超过八年以上结题的国家社科项目没有进行统计和分析。

最后,国家社科基金的学术研究成果方面。根据从CNKI获取到的由国家社科基金项目资助的学术论文,本专著主要从项目类型的学术论文产出数量分布、项目学术论文的类型分布、项目论文发表的期刊分布、项目论文的作者数量分布、项目论文的高频作者分布、项目论文的机构分布、项目论文的高频机构分布、项目的高被引论文和学术研究热点等几个方面进行了细致而全面的统计和分析。在统计论文机构的过程中,论文机构被设定为二级单位机构,比如,在统计高校这一类的机构时,二级单位设定到院或系。

在上述国家社科基金统计和分析的整体框架基础上,基于统计的方法、内容分析的方法和比较的方法,本专著虽然从历时的视角上对国家社科基金及其学术研究成果进行了多角度、多层面的呈现,但由于国家社科基金项目的数据和学术成果的论文数据均来自互联网,并且在确定统计和分析的具体内容方面均是从我们自己的研究兴趣点着手的,所以势必会造成统计和分析的不全、不当和不妥之处,敬请各学科的研究者、项目的管理者 and 对该研究感兴趣的读者谅解、批评和指正,我们定将逐一改正并万分感激。

王东波

2015 年 12 月

目 录

前言

1 引言	(1)
2 马列·科社	(11)
3 党史·党建	(28)
4 哲学	(43)
5 理论经济	(59)
6 应用经济	(75)
7 政治学	(92)
8 社会学	(109)
9 法学	(126)
10 国际问题研究	(141)
11 中国历史	(156)
12 世界历史	(172)
13 考古学	(188)
14 民族学	(204)
15 宗教学	(217)
16 中国文学	(233)
17 外国文学	(252)
18 语言学	(268)
19 新闻学与传播学	(284)
20 人口学	(300)
21 统计学	(316)
22 图书馆、情报与文献学	(331)
23 体育学	(348)
参考文献	(365)
后记	(370)

1 引言

自 1991 年全国哲学社会科学规划办公室成立以来,围绕着人文社会科学的马克思主义·科学社会主义,党史·党建,哲学,理论经济,应用经济,政治学,社会学,法学,国际问题研究,中国历史,世界历史,考古学,民族问题研究,宗教学,中国文学,外国文学,语言学,新闻学与传播学,图书馆·情报与文献学,人口学,统计学,体育学,管理学,教育学,艺术学,军事学 26 个学科,通过重大项目、年度项目、特别委托项目、后期资助项目、西部项目、中华学术外译项目 6 类主体项目,中国的人文社会科学研究得到了健康、有序和快速的发展。近年来,随着国家社科基金项目和经费的不断增长,对社科基金项目管理的科学化提出了新的要求和挑战。截至 2015 年 4 月,国家社科基金项目累计立项已达到 47058 项,如何对这一时间跨度已达 25 年的国家级项目从立项基本信息到结项基本信息再到项目的成果进行多层次、全方位和历时的评价,逐步被提上了议事日程。在这一大的背景下,通过获取每一学科的国家社科项目及发表的学术论文,本专著从国家社科项目的年度分布、项目类型分布、项目负责人、负责人职称分布、承担单位、标题的字(词)长度与高频词、项目结项时间跨度、专著出版机构、项目评定结果、项目类型发表论文数量、论文类型及期刊分布、作者数量及高频作者、机构数量及高频机构、研究热点统计和高频论文分析等几个方面对 25 年来的国家社科进行了统计和分析。

1.1 国家社科基金的相关研究

从提高项目管理效率的角度,East(1964)针对基于美国自然科学基金所产生的学术文献的集约化管理问题进行了相应的论述。通过系统整理和归纳信息学科的相关研究情况,Resnikoff(1980)完成了对美国自然科学基金所资助的信息学科项目的统计与分析。基于 1945—1975 这一时间跨度下的美国自然科学基金项目,Harden(2003)回顾了生物学研究的整体情况,并对研究的主要关注点进行了相应的分析。根据美国自然科学基金项目有关陶瓷学科的经费资助情况,Madsen 和 White(2014)对瓷器学科的工程性与教育问题的投资情况进行了系统的论述。以环境生物学的院系为单位,Watts 等(2015)对美国自然科学基金在促进环境生物学的发展中所发挥的作用进行了多方面的分析。从美国国家自然科学基金所资助的项目对美国民众科学素养影响的角度,Roos(2014)对时间跨度为 5 年的美国自然科学基金进行了多角度的分析。基于 2000—2011 年间的社会学、行为学和经济学的美国自然科学基金项目,Nichols(2014)通过主题模型对学科间的跨学科性进行了探究。从科学学和创新研究的角度,Hall(2014)对美国自然科学基金的研究机构及其相关研究进行了系统的分析。基于 360 个生物、工程、数学和物理学科的美国自然科学基金项目的摘要内容,Kamenetzky(2013)对科学基金的影响力标准进行了计量性的研究。Mata 和 Scheiding(2012)对社会科学在 20 世纪 70 年代末和 80 年代初兴起的原因进

行了多方面的探究。Gorman(2011)对美国科学基金中所涉及的科学、技术和社会三个方面的关系进行了探究。根据美国科学基金的研究报告,Simpson(2011)针对复杂的工程系统提出了改进的跨学科设计优化方案(MDO)。Jeffrey(2009)从资金投入和研究的需求角度对科学基金进行了分析。从美国科学基金资助的角度,Colón(2008)对美国化学生物学的研究情况进行了历时的统计和分析。基于国家科学基金,Ben等(2006)对创新力支撑工具这一研究专题进行了系统的分析。

陈媛媛、周鑫(2015)对国家社科基金数据库1994—2014年涉及“图书馆、情报与文献学”的项目数据进行了计量分析。通过文献计量学方法,许新军(2007)对十余年来获得国家社科基金资助的“图书馆、情报与文献学”项目的人员,从地域分布、单位分布、核心研究者等方面进行分析研究。以2006—2011年国家社科基金民族研究立项材料为样本,利用SPSS15.0和ROSTNAT软件,方付建(2012)对立项数目、机构、学科、议题、成果形式、研究群体等进行频数和社会网络分析。姚嘉慧等(2015)对1994—2012年国家社科基金项目中国图书馆、情报与文献学学科分类立项项目的整体情况进行了统计分析。在文献资料法、调查访问法、网络信息法和数理统计法等方法的基础上,金曼、李锋(2015)对2009—2013年江苏省体育学社科基金项目进行分析研究。谢征对1991—2013年国家社科基金结项的旅游类项目学术专著成果的出版情况进行了统计分析。黄华伟、傅秀兰(2012)以2001—2010年社科院系统国家社科基金项目立项情况为研究对象,分析了项目类别、年度分布、单位分布、学科分布,并与前期进行了比较。徐俊、风笑天(2014)以2004—2013年国家社科基金社会学年度立项数据为样本,使用SPSS20.0软件对近十年来社会学立项类别、机构、地区、单位等维度进行了分布特征分析。石磊等(2014)以2009—2013年国家社科基金“新闻学与传播学”立项项目为主要参照,从中归纳整理出了新闻学与传播学研究的学术热点。余春燕(2014)通过对2009—2013年国家社会科学基金资助文献学项目的统计分析,揭示了当前我国文献学研究的发展概况和研究热点。何小贞、廉超(2014)以2003—2013年我国31个省、直辖市、自治区的国家社科基金立项项目为研究对象,采用空间差异及演变特征测度指标,对我国国情类国家社科基金项目立项的空间差异进行分析。采用文献计量法,王海宏、王健(2014)以国家社科基金体育学(2009—2013)项目数据为研究对象,从立项类型、立项数量、立项单位、所属系统和课题指南与研究领域进行了数理统计。秦嘉杭(2013)对近三年(2011—2013年)国家社会科学基金在图书馆、情报与文献学类立项资助项目的关键词进行了归类与分析。王晴(2013)通过分析2012年度国家社科基金中国情档领域的项目名称,并与项目课题指南进行对比,归纳总结出新一年度图书馆学的研究热点和特征分布。张永汀(2013)采用文献计量学方法对1993—2012年期间我国政治学研究领域的国家社科基金项目情况进行了统计,对立项时间、立项数量、立项类型、地区分布、核心研究者、所属单位、立项成果形式等外部信息进行了全面系统的分析。闫洁(2013)选取近十年外国文学课题立项项目作为研究对象,通过项目类型、单位和地区分布等多维度的历时统计分析,同时通过自建2003—2012年国家社科基金外国文学立项课题项目名称语料库,从研究国家、文体和类型三方面对其立项情况进行实证的索引统计研究。石磊等(2013)以2009—2012年国家社科基金立项项目中“艺术学”和“新闻传播学”为择取范畴,遴选出影视学论题73项,对其进行整理与分析。周霞(2013)对2002—2011年期间的国家社科项目的整体情况进行了系统的统计与分析。王晓丽等(2015)对“九五”以来甘肃省国家社科基金的立项数量、单位分

布、学科结构及核心负责人等进行了多角度的统计与分析。唐雅琳(2015)统计了2011—2013年国家社会科学基金情报学立项课题的总体数量,对立项课题数量、立项类别年度分布、课题地域分布、立项人员机构分布、预期成果类型分布、立项主题等进行了详细的计量分析。基于2000—2012年的国家社会科学基金出版研究立项项目,孙艳、曾润喜(2013)从立项项目数量、项目类别、立项单位、项目负责人的职称、学科分类和立项主题6个方面对出版研究的发展状况进行了回顾和分析。通过对国家社科基金对学术期刊资助情况的统计分析,曾梁羚(2013)揭示了我国未来学术期刊的发展态势。张永汀(2013)用文献计量学的方法对1993—2012年期间我国政治学研究领域的国家社科基金项目情况进行了统计。陈梅芬(2012)对1999—2011年期间我国语言学研究领域的国家社科基金项目情况进行了统计。杨光、白玉(2013)通过对2008—2012年间国家社会科学基金(以下简称国家社科基金)立项项目的数量、类型、研究成果、地区分布与主题进行统计和分析,从宏观上反映了我国图书情报与文献学研究的轨迹、现状及发展趋势。徐芝兰(2012)以2000—2012年间安徽省国家社会科学立项项目为样本,通过项目的年度、学科、类别、机构、选题等情况的统计,实证考察安徽省学术生产的格局。刘婧(2012)以国家社科基金艺术学项目中戏曲学立项情况为研究对象,从戏曲学在艺术学里总体立项比例、高校与科研单位立项比例、核心主持人、课题地域分布等方面对此国家级课题进行分析研究。李秋杰(2012)对2001—2010年10年间图书情报学领域批准立项的国家社科基金项目的关键词进行了计量分析。黄华伟(2012)以“十一五”期间高校国家社科基金项目立项情况为研究对象,分析项目类别、地区分布、学校分布、学科分布,并与“十五”状况进行比较。黄建伟(2012)应用统计分析法,对近十年相关国家社科基金项目及其成果的具体信息进行了统计分析。姚颖靖、彭辉(2012)运用文献计量学方法对上海市1993—2011年国家社科基金项目的立项状况进行了统计分析。刘雪梅(2012)根据11年来(1999—2009)国家社科基金中新闻学与传播学类立项资助项目的数据,分析了相关项目的表面特征以及走势所显现的含义。黄华伟(2012)以近十年我国33所农林高校获国家社科基金项目立项情况为研究对象,统计分析项目的立项时间、立项类型、立项单位、学科分布、核心主持人等。董明(2010)以1993—2010年期间国家社会科学基金项目党史党建学科的立项数据为研究对象,分析党史党建学科项目总体数量、年度分布、项目类型、项目主持单位系统、项目研究内容及项目预期成果形式。王晓华、董明(2010)以1993—2010年期间国家社会科学基金项目考古学学科的立项数据为研究对象,分析考古学学科项目总体数量年度分布、项目类型、项目主持单位系统、项目研究内容及项目预期成果形式。杜芳娟、朱竑(2010)通过1993—2009年中国国家社会科学基金对“民族文化”资助项目进行了分析。李振(2010)通过对1993—2008年国家社会科学基金政治学类立项项目的统计分析,从立项数目、立项主题、项目成果形式和立项机构分布等方面呈现了中国政治学现阶段的发展状况。周志峰(2009)对图书情报学1994—2008年国家社科基金项目类型、立项时间、立项单位及所属系统、地区、项目负责人等形式特征和主题内容特征进行了统计与分析。伍玉伟(2009)对近十年我国档案学国家社科基金立项项目进行了统计与分析。彭云、王伦安(2009)以近五年四川省获得的国家社科基金项目情况为研究对象,运用统计学方法,分析项目依托单位、学科领域、项目类别、成果形式、年度分布等。王玉琴(2009)通过对国家社科基金项目宗教学科课题的历年资助项目数据统计,并结合课题指南,分析其所立项目的特点。纪蔚蔚(2009)以1998—2007年《中文核心期刊要目总览》“科学、科学研究”类

期刊中的国家基金论文为统计数据源,运用文献计量学的原理和方法,对中国科学学国家基金论文生产力发展水平进行计量学研究,在这一研究的基础上,王东波(2015)基于1998—2012年间的科学学的国家社科项目对这一学科下的社科项目及其学术成果进行了更进一步的探究。鄂丽君(2009)从项目年度分布及类型、项目的地域分布、高校分布、承担课题的高校及二级机构、项目负责人、项目内容等方面对高校系统承担国家社科基金该类项目的研究现状进行了分析,并统计了自1994年至2008年10月图情国家社科基金论文的产出情况。陈秋玲(2009)对1993—2008年期间我国外国文学研究领域的国家社科基金项目情况进行了统计。杨荔媛、朱庆华(2008)以“图书馆、情报与档案管理”一级学科2000年以来的国家社会科学基金和自然科学基金项目为研究对象,从项目数量、项目类型、负责人及其单位、项目主题等方面进行了全面系统的比较分析。凤元杰、范全青(2008)对浙江省1993—2006年国家社科基金立项的项目从学科、地区、行业系统等方面进行综合分析。韩兆洲、安康(2008)利用统计学国家社科基金项目的数据,对我国统计学发展导向的问题进行了探讨。蔡尚伟、刘锐(2008)对1993—2007年间新闻学与传播学国家社科基金立项项目进行了综合分析。吕国光(2008)运用文献计量学的方法对湖北省1993—2007年国家社科基金项目立项状况进行了统计分析。张俊瑞、王鹏(2008)通过对1993—2007年15年间国家社科基金应用经济学科资助项目的统计分析与研究,发现了社科基金在资助学科、研究领域、选题上的一些规律。汪静媛、赵良英(2008)对1993—2007年近15年间国家社科基金资助经济学项目的类型分布、时间分布、隶属系统分布、地区分布、学校分布、主持人分布等进行了统计分析。吕国光(2008)对1993—2007年国家社科基金22个学科获准立项的13000余项课题进行了多角度的统计和分析。邵伟德、王守钧(2003)对1997—2002年国家社会科学基金(体育学)立项项目进行了统计分析,结果表明立项数量和资助力度逐年上升。田雨普和张文静(2006)采用文献资料调研和归纳等研究方法,对1997—2005年国家社会科学基金项目体育学研究课题进行分析。王守炳(2003)通过自制研究工具对党校系统承担项目数量、学科分布、单位分布、研究内容等进行了介绍和分析。冯昆思(2005)介绍和分析了有关民族问题项目的立项数量、类型、成果形式、主持人所属地区以及项目选题内容范围、频度、动态及倾向热点。凤元杰、范全青(2005)发现1993—2004年国家社科基金立项项目的学科与地区分布极不平衡。徐元君(2004)对1997—2003年国家社科基金体育学研究项目申报立项状况进行了统计分析与研究。刘君(2008)对1993—2007年15年间贵州省获准立项的国家社科基金资助项目,从立项年代分布、项目类型分布、项目学科分布、立项机构分布、项目内容特征等方面进行了全面的统计和分析。唐雪琼等(2008)通过对1993—2007年批准立项的国家社科基金有关女性项目进行归类统计。贺显斌(2005)了解过去十多年的国家社科基金立项情况。王红玲、张齐增(2005)对图书馆、情报与文献学学科立项基本情况和立项研究内容进行了简要分析。陈雪梅(2005)对近十年获国家社科基金资助的“图书馆、情报与文献学”研究课题做了穷尽性的分析。金霞等(2015)运用文献资料法、数理统计法和逻辑分析法等研究方法对2010—2014年国家社科基金体育学一般项目的立项区域分布与内容等进行了统计分析。徐俊等(2014)基于178篇国家社会科学基金立项研究综述的文献分析,回顾相关研究对于国家社会科学基金立项的分析论域和研究特点,评价其存在的不足。田时中等(2014)对国家社会科学基金资助的管理学研究项目的立项数量、立项区域、立项单位、预期研究成果等进行了统计分析。李英(2014)以图书情报与档案管理学科2009—2013

年的国家社会科学基金和国家自然科学基金项目立项情况为研究对象,从立项数量、立项类型、立项单位及其所属行业、项目负责人以及项目的主题内容特征等方面进行系统分析比较。崔雁黎(2013)通过对2003—2012年间国家社会科学基金立项项目的数量、类型、主题、立项单位类型、地域和预期成果进行统计和分析,揭示我国图书情报文献学领域研究的现状、热点与发展趋势。张佳慧(2013)以2008—2012年国家社会科学基金低碳研究立项项目为样本,从立项数量时间分布、立项类别、所在地域与机构、所述学科分布、项目预期成果形式和立项主题展开研究,从而回顾我国低碳的发展状况。高毅(2013)以国家社会科学基金法学项目为样本,从项目数量、类别、地域、预期成果、学科分布、研究群体等方面进行统计分析,展现当代中国法学学术生产格局。基于2010—2012年新闻学国家社科基金立项项目的论题选取,庞弘(2013)对新媒体及其效应、全球化背景下的媒介状况、传媒时代的法制与伦理、突发性公共危机的报道和处理、文化产业与传媒发展、受众及其舆论引导等研究热点进行了分析。刘泽梅(2013)对2006—2012年度国家社科基金项目体育类立项项目进行统计,从年份、地域、所在单位、项目类别、学科分类以及预期成果6个方面进行了分析。基于翻译研究的分类视角,采用内容分析法,年晓萍(2014)对近十年国家社科基金翻译立项进行分布特征的统计描述和主题内容的量化分析。通过对1993—2011年国家社会科学基金著作权问题研究立项项目的统计与分析,张文娟(2012)发现该类研究以青年项目和一般项目为主要资助类型、法学、图书馆·情报与文献学和民族问题研究是立项的主要学科。通过对近十年来国家社科基金公布的课题指南及获批的资助项目的文献数量特征和变化规律的描述,林艳丽、李坚(2012)探索了我国社会保障领域研究的动态、特点、趋势与发展的轨迹。张晓阳、窦美玉(2012)对“图书馆、情报和文献学”国家社科基金1997—2008年208个立项结题项目进行统计分析,揭示该领域项目的结项率、结项时限、结项成果、项目鉴定等级等方面的特征。刘文云、耿庆军、李勇(2007)对1994—2006年的294项图书情报的国家社科项目进行从基本信息到主题内容的多角度统计和分析。

1.2 研究的整体框架

本专著的整体框架由数据的抓取、数据的清洗、数据的统计和数据的分析4个主要部分构成,具体4个组成部分的内容如下:

首先,面向网络抓取国家社科项目基金的基本信息以及相应学术论文的信息。抓取的数据主要由两部分组成,一部分为面向“国家社科基金项目数据库”获取国家社科的基本信息,该信息主要包括如下内容:项目批准号、项目类别、学科分类、项目名称、立项时间、项目负责人、专业职务、工作单位、单位类别、所在省区市、所属系统、成果名称、成果形式、成果等级、结项时间、结项证书号、出版社、出版时间、作者、获奖情况,具体抓取的国家社科基金项目数据库页面如图1-1所示,本专著获取数据的时间为2015年5月10日,通过程序获取到了在这一时间之前的“国家社科基金项目数据库”中的所有社科项目。

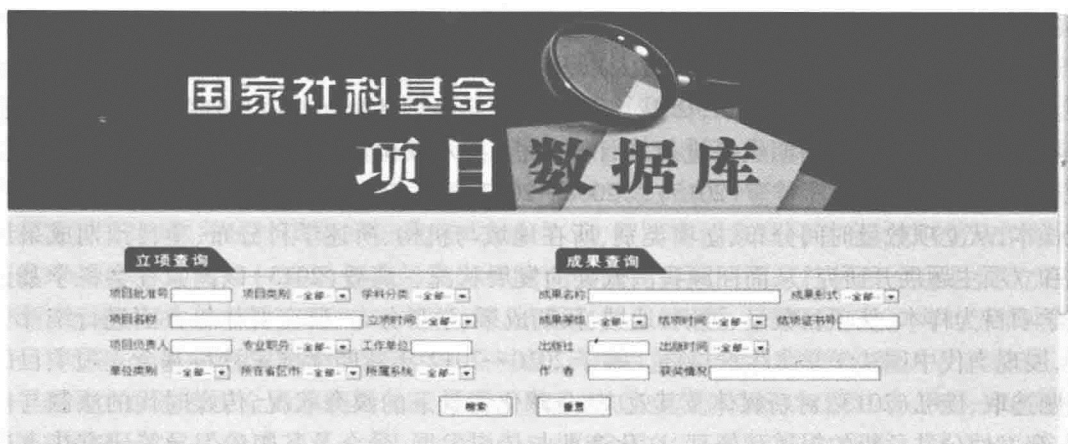


图 1-1 获取国家社科项目的国家社科基金项目数据库图例

另一部分为通过国家社科基金项目的名称或者编号,基于开发的专门抓取程序,从中国知识基础设施工程(CNKI)网络平台上获取该基金资助的学术论文的基本信息,具体的信息主要包括如下内容:学术论文的标题、学术论文的作者、学术论文作者的机构、发表学术论文的期刊、学术论文的关键词、学术论文的被引次数、学术论文被下载的次数等基本信息。抓取的 CNKI 的网络页面如图 1-2 所示,本专著抓取 CNKI 上的学术论文的时间跨度为 2015 年 7 月 1 日—2015 年 7 月 28 日,所获取的学术论文时间跨度为 1991—2015 年间的国家社科基金项目资助的所有论文。

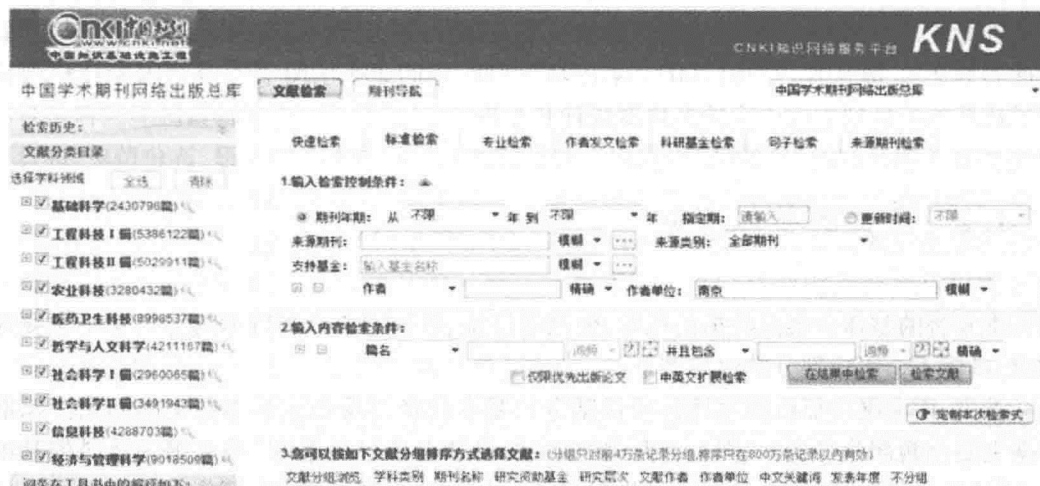


图 1-2 获取学术论文的 CNKI 图例

其次,完成对所抓取社科项目基本数据及所发表学术论文数据的清洗。该部分的工作主要由两部分构成,一方面由人工制定相应的清洗规则,比如高校名称变化表,另一方面在制定的规则基础上,通过设计相应的程序完成对数据的清洗。本部分主要清洗的对象包括国家社科项目单位的清洗、结项时间的规范、抓取学术论文的自动去重、资助项目名称的规范等。

再次,完成对国家社科项目和所发表学术论文的统计。数据统计主要由两部分构成,一

部分为国家社科基本信息的统计,从本统计报告研究的角度出发,主要统计了某一学科社科项目的年度分布、社科类型分布、项目负责人及其职称分布、项目承担单位分布、项目结项时间分布、项目专著出版情况,另一部分主要统计了不同类型项目发表学术论文的数量情况、学术论文的类型、学术论文的期刊分布、作者数量分布、高频作者情况、发表学术论文机构分布、高频机构情况、学术论文的研究热点情况、高被引论文分布情况。

最后,基于统计数据完成了对国家社科项目及所发表学术成果的分析。在分析的过程中,主要通过比较的方法、内容分析的方法、时间序列方法对国家社科项目的基本信息与所发表的成果进行了多角度的统计分析。具体研究的整体框架见图 1-3。

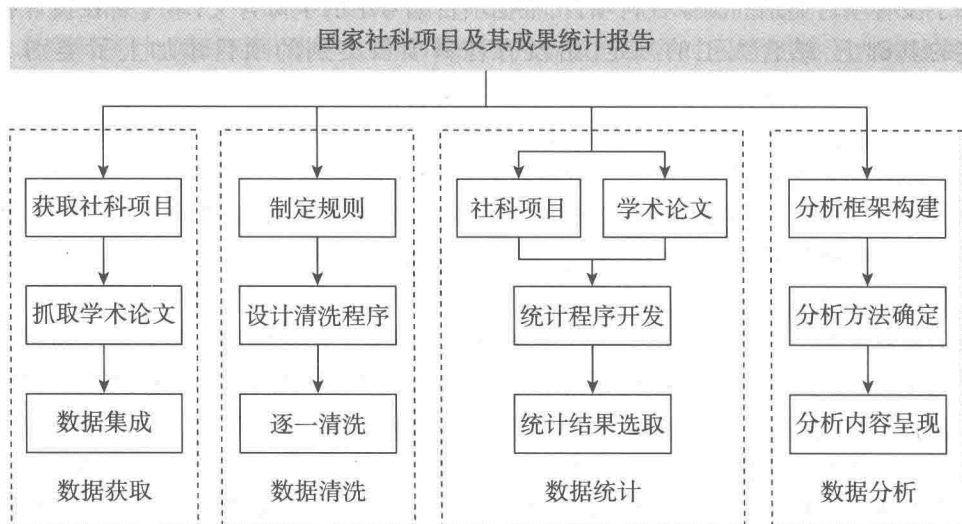


图 1-3 国家社科项目及其成果统计报告研究整体框架

1.3 相关数据的说明

(1) 关于国家社科项目基本数据的说明

虽然国家社科项目在 1986 年就设置了,但全国哲学社会科学规划办公室是由中央于 1991 年在全国哲学社会科学规划领导小组下设立的。规划办成立后,国家社科项目的申请、立项、结项形成了统一的管理规范,同时在已有相关国家社科项目基本信息的基础上开发了国家社科基金项目检索平台。从数据的规范性、完整性角度考虑,本专著统计的国家社科项目是从 1991 年开始的,因此对于相关社科项目的数据说明如下:

凡是立项时间在 1991 年之前的社科项目基本数据均不在本统计报告当中呈现,比如在抓取的数据中对于 1987 年立项的“教学心理研究与应用”和“中国儿童心理发展的特点与教育”这类项目。同时“国家社科基金项目数据库”中标注错误的数据记录也不在本统计报告中,比如“123456 重大项目 0 dasdasda 0000-01-01 dasd 0”“批准号 课题名称 0000-00-00 首席专家”。

对于在抓取的数据中凡是没有标注具体时间的项目,如果可以根据项目编号推断出具

体年份的,通过人工标注上具体的年份,比如对于“10BDJ026 一般项目 党史·党建 中国共产党全国代表大会制度研究 0000-00-00 孙功 副高级”这一条国家社科项目的信息,根据项目编号“10BDJ026”可以推断出该项目立项的时间为2010年。

对于没有标注学科类别的国家社科项目,本专著主要根据项目名称,结合搜索引擎获取到的项目负责人的信息,给社科项目标注相应的学科类别,以便于统计。在没有学科类别的项目中,重大社科项目相对校对,比如对于“09&ZD062 重大项目 中国特色社会主义司法制度研究 2010-01-30 陈桂明 正高级”,根据项目名称和项目负责人,判定该类项目为“法学类”。

对于没有项目类别的国家社科项目,根据项目编号中的字母含义,本专著在设计相应软件判定的基础上,结合人工的判定,给没有社科项目类别的项目添加上了类别。比如“15AZD067 流域一体化治理的法律机制研究 0000-00-00 李爱年 正高级”这一个项目中,根据“AZD”推断出该项目为重大项目。

对于项目主持人的职称,本专著只统计了“国家社科基金项目数据库”中标识了项目主持人具体职称的项目,对于没有标注项目负责人职称的项目则不在统计之列。同时,本专著中所指的职称主要涵盖如下四个层级“初级、中级、副高级、正高级”,把凡是国家社科项目中标注为“教授、副教授、研究员、副研究员、讲师”等职称名称的统一归并到上述四个类别中。

对国家社科项目负责人的单位进行了统一规范。国家社科项目负责人所在的单位的规范是较为复杂的,主要有以下几个原因:单位名称变更,比如高校合并;具体二级单位的名称不一致;同一层次上的单位名称不统一。针对上述几种原因,本专著处理策略如下:首先,提取出所有国家社科项目负责人所在的单位名称,对单位名称进行频次统计,根据统计的结果共获取到了9990个社科项目负责人的单位。其次,对9990个单位,通过人工逐条进行处理,对于名称已变更的单位,以现在的名称为主,对于包含二级单位的机构名称进行统一的归类处理,只保留上一层架的单位名称,对于同一性质的单位,在名称上进行一致性归类处理,具体机构名称处理的规范样例见表1-1。最后,在上述人工整理的规则基础上,设计专门的程序完成对社科基金项目负责人机构的规范。

表1-1 国家社科项目负责人单位规范样例表

编号	规范前机构名称	规范后机构名称	机构规范的规则
1	中国社会科学院哲学研究所	中国社会科学院	去掉二级单位
2	中共北京市委党校	北京市委党校	统一单位名称
3	武汉大学经济与管理学院	武汉大学	去掉二级单位
4	徐州师范大学	江苏师范大学	归并到新的名称
5	徐州师范学院历史系	江苏师范大学	归并到新的名称并去掉二级单位名称
6	宁夏医学院	宁夏医科大学	归并到新的名称
7	中共青海省委党校党史党建教研室	青海省委党校	统一单位名称并去掉二级单位
8	江苏南京师范大学外文系	南京师范大学	统一单位名称并去掉二级单位

续表

编号	规范前机构名称	规范后机构名称	机构规范的规则
9	陕西省咸阳师专	咸阳师范学院	归并到新的名称
10	江西财经学院农业经济系	江西财经大学	归并到新的名称并去掉 二级单位名称

对项目结项的时间跨度进行了相应的处理。本专著对社科项目结项的时间跨度是通过课题结项时间点与立项时间点的差值计算出来的,计算单位为年。本专著报告只呈现了“国家社科基金项目数据库”中标识了这两个时间的项目,对于没有标注这两个时间的社科项目本文没有进行统计,同时对于计算出的时间跨度为负值的或者时间跨度过于大的社科项目,在专著分析报告中也进行了相应的删除。

对于项目结项的评定级别这一数据,本专著只统计了“国家社科基金项目数据库”中已经标注了具体结项级别的社科项目,对于没有标注结项级别的国家社科项目,本研究没有进行统计,具体的评定级别主要由“优秀、合格、良好”三种类型构成。

对于出版社科项目专著的出版机构,本专著也只统计了“国家社科基金项目数据库”中社科项目结项成果为专著并且提供了专著出版机构的社科项目,对于没有提供出版机构的国家社科项目,本专著没有进行统计和分析。

对于所统计的国家社科项目标题中含有的字和词的长度分布这一数据,在计算字和词的长度过程中,主要进行了如下处理:统计字和词的长度时把具体的数字比如“20”视为一个长度单位进行了计数;对国家社科项目的标题进行了自动分词,自动分词是调用“NLPIR - ICTCLAS2015”程序完成的,具体分词的样例如下:没有分词社科项目标题:“毛泽东研究中的历史虚无主义观点评析”,分词后的标题:“毛泽东/研究/中/的/历史/虚无主义/观点/评析/”。

(2) 关于国家社科项目资助的学术论文数据说明

对于不同项目类型的学术论文产出分布情况这一数据,如果某一篇文章学术论文是由多个项目资助的,这一篇文章学术论文则各在资助的项目下计数一次,同时在统计相应的成果过程中,如果是“国家社科基金特别委托项目”的成果则不在统计之列。对于学术论文发表期刊分布这一统计的数据,本专著只统计了 CNKI 所包含的期刊发表的学术论文,对于 CNKI 没有收录的期刊,本专著没有进行统计,比如《情报学报》这一期刊,CNKI 中没有收录,统计数据中就没有《情报学报》的相应统计数据。

对于学位论文的作者分布数据,一方面,根据一篇论文的作者数量,统计了学术论文的作者数量分布,具体统计的过程中是根据 CNKI 中对论文作者进行分开的“;”这一符号进行的,比如“刘文云;耿庆军;李勇;”这一作者序列有三个“;”则计数为3。另一方面,对于学术论文的高频作者分布这一数据,在统计的过程中没有区别第一作者、第二作者或者通讯作者,只要作者出现在学术论文中则被计数一次。

学位论文的机构分布数据,一方面,根据一篇学术论文中的机构数量,统计了学术论文的机构数量分布,具体统计的过程也是根据 CNKI 中对论文作者所在机构进行分开的“;”这一符号进行的,比如“南京农业大学信息科学技术学院;南京大学信息管理学院;南京大学双

语词典研究中心;”这一机构组合中有三个“;”则计数为3。另一方面,对于学术论文的高频机构分布这一数据,在统计的过程中只要机构出现一次就为该机构计算一次;由于发表学术论文的机构基本上是由二级单位构成的,所以在统计发表学术论文的高频机构的过程中只保留了二级单位的机构,没有呈现一级高频机构单位的分布情况。

对于高频关键词这一数据,本专著根据从CNKI中获取到的国家社科项目资助的论文中的关键词序列,统计了每一关键词的出现次数,根据分析的需要选取了前50的高频词。在分析的过程中,对于关键词没有进行筛选,主要是根据出现频次列举在了每一学科的分析当中。

对于“管理学”“军事学”“艺术学”和“教育学”这四个学科,由于“国家社科基金项目数据库”中关于“管理学”的社科项目只提供了2004—2010年间的社科项目,整体数据量缺失较大,所以在本统计报告中没有统计“管理学”这一学科。“军事学”和“艺术学”的社科项目整体数量较少,缺乏统计的价值,所以也不包括在本报告之中。“教育学”这一学科的2013年之前的社科项目在“国家社科基金项目数据库”中均没有社科项目编号,这一情况既不方便统计社科项目本身也不便于获取社科项目所发表的学术论文,因此本报告也没有对“教育学”进行统计和分析。

1.4 小结

首先,本章系统地梳理了国内外有关国家资助项目的研究情况,主要是围绕着有关国家社科项目的相关研究进行了多角度、多层面的文献综述,从微观上有关于某一学科的国家社科项目的情况,从宏观上也有关于某一时间跨度下的国家社科项目的研究情况。其次,本章给出了研究的整体框架,本框架主要涵盖了数据的获取、数据的清洗、数据的统计和数据的分析四个方面,重点对数据的获取和数据的清洗进行了详细的介绍。最后,本章对正本统计报告中所使用的数据进行了详细的介绍,主要是使用了哪些数据、剔除了哪些数据、对哪些数据进行了细致的加工等。