

越是民族的就越是世界的,越是民族的也越应走向世界。在全球化、信息化的今天,作为少数民族文化的载体,少数民族文字理当如是。

敲开少数民族文字信息化大门

■本报记者 李芸

教计算机“识字”

少数民族文字信息化,最基础的工作是信息的“输入”。输入,最常见的方式是人工用键盘输入。可面对浩如烟海的图书与典籍,要把这些优秀的少数民族文化逐字逐句地手工输入计算机,简直是不可能完成的任务。

这就有了扫描。但扫描只是将文档图像输入计算机,充其量只是输入的初级阶段,因为语言文字及文档通过扫描仪输入计算机中,在计算机看来,还只是一张照片,是一堆不知所云的图像点阵,计算机不能认识它,也不可能基于内容的全文检索、查询浏览等功能。

于是,把图像点阵转化为计算机可阅读并检索的文字编码,让计算机“识字”,就成了最关键的问题。

只认识“0”和“1”的计算机是如何“识字”的?近日,本报记者采访了计算机的“识字师傅”之一、清华大学教授丁晓青。

丁晓青与她的团队——清华大学电子工程系智能图文信息处理研究室的师生们,最先挑战了汉字。

那是20世纪80年代中期,一开始使用的方法是拆笔画识别算法。

因为汉字是一笔一画写就的,顺势思维便是让计算机一笔一画提取文字信息。“可汉字太过复杂,很多字笔划复杂,识别错误率太高。”丁晓青说。

而彼时英文识别采取的是投影加笔划识别,世界范围内少有可供汉字识别借鉴的方法。

为什么不将汉字作为一个整体来识别呢?1989年,丁晓青研究组独辟蹊径,采用图像识别的方法。他们让计算机通过图像来进行“全局性”认字,就像我们认字往往凭借视觉的感知一样,让计算机通过对文字整体结构的判断来认字。

化零为整,错误率反而降低了,高效且实用。

丁晓青说:“后来我们利用图像识别的方法去识别其他文字,如认英文、日文、韩文等,也完全取得成功。理论上,图像识别的方法适用于所有文字。应该说,我们找到了解决计算机识字的普适性方法。”

普适性的钥匙

解决汉字难题后,下一步就是少数民族文字。

我国是统一的多民族国家,但直到21世纪初,与拉丁文、汉字相比,蒙古、藏、维吾尔、哈萨克、柯尔克孜等民族文字的识别研究几近空白,严重影响了文化信息化的发展。

虽然手握图像识别这个具有普适性的工具,但少数民族文字识别的研究并非坦途一片。由于各民族文字历史渊源、语言规范不同,构字法、构词法不同,语法、书写习惯(从左向右,或从右向左,或从上向下)也不同,民族文字与方块汉字有本质不同。

丁晓青说:“譬如汉字横平竖直,少数民族文字字形不规则,弧线笔画多;汉字是方块字,少数民族文字则字形无定长、无定宽、变化大,有不同方向的连笔书写,字符边界也不确定。”

丁晓青选择最重要、最具代表性的蒙藏维哈柯文字为突破口。清华大学于1999年开始藏文识别研究,与合作单位西北民族大学在2003年完成了藏文(混排汉英)文档识别系统;于2002年开始了维吾尔文、哈萨克文、柯尔克孜文以及阿拉伯文识别研究,与合作单位新疆大学在2004年完成了印刷体维吾尔文、哈萨克文、柯尔克孜文、阿拉伯文(混排汉英)文档识别系统;于2005年开始蒙古文识别的研究,与合作单位内蒙



古大学、内蒙古师范大学在2007年完成了多体蒙古文(混排汉英)印刷文档识别系统。

通过对实际文档的识别性能的测试表明,系统对各种文字单字识别率均达99%以上,实际民族文字(含汉英混排)文档识别率达97%,均处于世界领先水平。

十多年下来,丁晓青这位师傅已远远不如徒弟了,她笑言如今是“计算机认识的字,我不认识”了。

在该技术上推出的“蒙藏维哈柯朝主要民族文字汉英混排文档综合识别跨语种理解系统”,为促进中华各民族文化信息化及交流提供了重要的数字化工具,并广泛应用于印刷体维吾尔文、哈萨克文、柯尔克孜文、阿拉伯文(混排汉英)文档识别系统;于2005年开始蒙古文识别的研究,与合作单位内蒙

走向“沟通无极限”

“解决大量民族文档自动输入计算机的难题,是实现民族历史和当今文化信息的查询检索等信息化处理首先必须解决的问题。”丁晓青说,“但这项研究又是‘打地基’的工作,往往看不到太多的经济效益。国家自然科学基金委员会却一直在这项事业中给予稳定持续的支持。”

据悉,国家自然科学基金委员会自成立以来,以青年、地区、面上、主任基金、重点项目的方式,对少数民族文字信息的处理与识别研究资助了近百个项目,资助经费达1500余万元。国内清华大学、西北民族大学、内蒙古大学、新疆大学、西藏大学等单位的研究人员,在基金的资助下,取得了显著的成果。

虽然多民族文字识别取得了突出成果,但仅仅是一个开始,在丁晓

青看来,民族文字信息化还有大量的工作需要去做。

丁晓青和她的团队现今正在做少数民族文字的“识别理解”工作,“例如,可以将藏文文档经过识别,变成计算机可以阅读的文档,这样就有可能利用字典查询等工具,将其内容转换为汉字加以理解,或反之,将汉文文档转换为藏文或蒙文文档表示。我们把这称为‘跨语种识别理解’,是通过文档识别及机器跨语种理解等智能信息手段,实现文字‘查字典’的功能。”

“我希望用这种方式实现各不同语种的民族语言及文档信息之间的的种转换,而使民族文字语言真正实现‘沟通无极限’还需要进行大量如‘机器翻译’等的研究开发工作。这对推动民族信息化,促进无障碍的民族文化交流,对维护国家统一和民族团结也是具有重要战略意义的。”丁晓青说。

科学基金在高校

中国海洋大学:

咬住特色不放松 交叉融合促发展

■本报记者 廖洋

通讯员 魏俊峰 刘彬

坐落于魅力海滨城市——青岛的中国海洋大学,是一所海洋和水产学科为显著特色,国家“985工程”和“211工程”重点建设的教育部直属综合性大学,国家自然科学基金一直是学校科研经费的四大支柱之一。

据统计,自国家自然科学基金设立至今,该校共获基金资助943项,34600余万元。其中,国家创新研究群体1项,国家自然科学基金重大基金1项,国家杰出青年基金获得者13人。

中国海洋大学副校长闫菊告诉记者,国家自然科学基金为学校凝聚和稳定有创新能力的研究队伍,发展一批具有国际影响力的研究方向,促进学科交叉研究,增强源头创新和争取承担国家各项重大科研项目的能力等科学研究实力提高起到了至关重要的作用。

成绩源于“特色”的不懈坚持

据统计,仅2006年-2011年,该校基金经费数和项目数年均分别增长39.3%和21.1%。在此期间学校累计获科学基金项目513项,经费25729万元,其中创新群体1项,国家杰出青年基金项目4人,重点基金20项,重大国际合作项目4项。

对此,闫菊表示:“科技奖励是学校科技水平的集中体现,国家各级各类科技计划的长期稳定支持,为科技奖励提供了基础和保障。学校以第一完成单位获得国家技术发明一等奖1项、国家技术发明二等奖2项、国家科技进步二等奖5项,基金在其中起到了显著的促进作用。”

然而,成绩的取得离不开特色的坚持和不懈的努力。中国海洋大学科学技术处副处长罗轶介绍,学校十分

重视科学基金的特色管理,按照“强化发展特色、协调发展综合、以特色带动综合、以综合强化特色”的发展思路,充分发挥海洋、水产学科综合的优势,引导教师在海洋、水产方向上做深、做精,努力做到国内领先,国际知名。

值得一提的是,该校在自然科学基金地球科学部海洋学科,近十年获资助列全国科研单位第一。“通过‘方向—平台—项目—人才’四位一体的科技工作模式,以及‘十一五’期间的累积和发展,我们获项目资助呈良性发展态势,多年覆盖基金委所有8个学部。”闫菊表示。

管理提高质量 服务保驾护航

众所周知,科学研究实力是一所高校整体实力的重要组成部分,广大科研工作者更是高校科技发展的中坚力量。

“因此,科技处作为学校科研管理部门,更为重要的是要做好服务工作,通过管理水平的提高,切实做好为科学家们的服务工作,解除他们的后顾之忧。”罗轶说。

中国海洋大学一直把科学基金管理从源头即申请时抓起,在每个项目申请过程中,通过帮助教师分析当年形势、写好申请书,并组织专家单独进行多轮预评审,年初定期召开申请动员大会等多种形式,以周密细致的工作保证了项目的质量。

据悉,“十一五”期间,中国海洋大学面上项目平均资助率为28.5%,高于全国水平近10个百分点。

“申请工作仅仅是基金管理的开始,更重要的是对资助项目的过程管理。”罗轶告诉记者,对于资助项目,学校积极为项目负责人和学术骨干提供专业化指导和服务,在项目执行过程中,定期组织专家检查项目执行情况,

帮助科研人员解决实际困难,为项目顺利进行保驾护航。同时,学校也十分重视科学基金项目的绩效管理,如倡导教师发表高水平的学术论文,并以此作为晋升高职称的条件。

宋微波教授便是在此基础上成长起来的优秀代表。20多年来,在以国家自然科学基金项目为主的资金扶持下,他学以致用,长有发挥,并逐步成长为国际原生动物界具有重要影响力的科学家。

宋微波所从事的海洋原生动物学研究在国内是一个十分薄弱的领域,他1989年初回国时,正处于国内科研大环境的低谷期。在百般艰难中,国家自然科学基金及时纾解了他经济上的困顿,并连续多年来为其提供了全面的支持;他自1990年承担第一个基金项目起,至今先后承担6项面上项目、2项重点项目、1项国家杰出青年科学基金项目、1项青年基金、多项中英、中德国际合作项目。这些工作先后在国际主流刊物上形成了280余篇学术论文,出版了专著与专集4部。成果分别获得国家自然科学二等奖及教育部、国家海洋局自然科学一等奖。

如今,他创建和领导的中国海洋大学原生动物学研究室已成为公认的海洋原生动物学研究中心,其学术成果和水平被国际同行在不同场合下誉为world-leading, world authority,并在2011年当选国际原生动物学会5人执委之一。

“自然基金风雨二十五载,蕴含了无数基金人的心血和汗水。在未来‘十二五’的新征程上,学校将坚持国家自然科学基金委员会提出的尊重科学、鼓励探索、平等宽容、激励创新、公正透明、民主和谐的价值取向,继续强化发展特色,提高管理质量和服务水平,进一步做好我校的基金管理科研工作。”闫菊表示。

指南

2012年申请须知(1)

依托单位和申请人在申请2012年度国家自然科学基金项目时,应当遵守下列规定。

关于申请人条件

1. 依托单位的科学技术人员作为申请人申请国家自然科学基金项目,应当符合《国家自然科学基金条例》(简称《条例》)第十条第一款规定的条件:具有承担基础课题研究或其他从事基础研究的经历;具有高级专业技术职务(职称)或者具有博士学位,或者有2名与其研究领域相同、具有高级专业技术职务(职称)的科学技术人员推荐。部分类型项目在此基础上对申请人的条件还有特殊要求。

2. 从事基础研究的科学技术人员,具备《条例》第十条第一款规定的条件,无工作单位或者所在单位不是依托单位,经与在自然科学基金委注册的依托单位协商,并取得该依托单位的同意,可以申请面上项目、青年科学基金项目,不得申请其他类型项目。

该类人员申请项目时,应当在申请书个人简历部分详细介绍本人以往研究工作及现工作单位情况,并提供依托单位同意本人申请项目的证明,作为附件随纸质申请书一并报送。

3. 正在攻读研究生学位的人员(自然科学基金接收申请截止日期前尚未获得学位)不得作为申请人申请各类项目,但在在职人员经过导师同意可以通过受聘单位申请部分类型项目,同时应当单独提供导师同意其中项目并由导师签字的函件,说明申请项目与其学位论文的关系,承担项目后的工作时间和条件保证等,作为附件随纸质申请书一并报送。

在攻读研究生学位的人员可以申请的项目类型包括:面上项目、青年科学基金项目、地区科学基金项目及部分联合基金项目(特殊说明的除外),但在攻读硕士学位期间,不得申请青年科学基金项目。

4. 正在博士后工作站内从事研究的科学技术人员申请国家自然科学基金项目,需要由依托单位提供书面承诺,保证在获得项目资助后延长其在博士后工作站的期限至项目资助期满;或者是出站后继续留在依托单位从事科学研究。否则,自然科学基金委不予受理在站博士后人员的项目申请。

关于申请书撰写要求

(一)申请人在撰写申请书之前,应当认真阅读《条例》、2012年度《指南》、相关类型项目管理办法和有关受理申请的通知、通告等文件。现行项目管理办法与《条例》和2012年度《指南》有冲突的,以《条例》和2012年度《指南》为准。

(二)申请书应当由申请人本人按照撰写提纲撰写,并注意在申请书中不得出现任何违反法律及有关保密规定的内容。申请人应当对所提交申请材料的真实性、合法性负责。

(三)根据所申请的项目类型,准确选择“资助类别”、“亚类说明”、“附注说明”等内容。要求“选择”的内容,只能在下拉菜单中选定,要求“填写”的内容,可以键入相应文字;有些项目“附注说明”需要严格按2012年《指南》相关要求填写。

(四)根据所申请的研究方向或研究领域,按照2012年《指南》所附的“国家自然科学基金申请代码”准确选择申请代码,特别注意。

1. 选择申请代码时,尽量选择到最后一级(6位或4位数字,重点项目和联合基金项目等特殊要求的除外)。

2. 申请人选择的申请代码1是国家自然科学基金委员会确定受理部门和遴选评审专家的依据,申请代码2作为补充。部分类型项目申请代码1或申请代码2需要选择指定的申请代码。

3. 申请代码首位字母为“L”、“J”的,属于专用申请代码,仅在申请特殊类型项目时可以选择。如申请代码首位为“L”的,仅用于申请NSFC—广东联合基金、NSFC—云南联合基金、NSFC—新疆联合基金等项目;首位为“J”的,仅用于申请国家基础科学人才培养基金、青少年科技活动、局(室)委托任务等特殊要求项目。如果在面上项目、青年科学基金项目、地区科学基金项目等类型项目申请时选择了以上的申请代码将不予接收。

4. 2012年部分学科领域拟试行“申请代码”、“研究方向”和“关键词”的规范化选择,包括物理学(D01及其下属申请代码)、电子学与信息系统(F01及其下属申请代码)和肿瘤学I(H1601至H1614申请代码)。

上述学科领域项目的申请人在填写申请书简表时,应参考“试点学科领域申请代码、研究方向和关键词一览表”准确选择“申请代码1”及其相应的“研究方向”和“关键词”内容。该一览表详见国家自然科学基金委网站(<http://www.nsf.gov.cn/>)“申请受理”栏目下的“特别关注”。

5. 申请人如对申请代码有疑问,请向相关部门咨询。

(摘自《2012年度国家自然科学基金项目指南》)

中国科学基金	
第26卷第1期 2012年1月	
目次	
凝聚合力 锐意进取 共创中国特色科学基金制发展新局面.....陈宜瑜 1	试谈做人做事做学问.....师昌绪 6
学科进展与展望	
全球气候变化与农业有害生物控制研究进展及热点——第66期“双清论坛”综述.....戈峰 马占鸿 胡琼波 罗晶 9	多尺度海气相互作用研究展望——第61期“双清论坛”综述.....苏纪兰 吴国雄 周天军 陈大可 王桂华 13
科学论坛	
试论博士后开展独立科学研究的能力和在科研工作中的定位.....曹凯 宿芬 孟祥利 李兰欣 17	地方高校基础研究现状及发展对策.....张琪 22
科研诚信建设应从教育抓起——科研诚信教育座谈会综述.....25	对科学基金评审制度的思考.....王瑞 28
成果简介	
重大国际合作项目实施成效与经验——2010年度化学科学部重大国际合作结题项目成果介绍.....黄宝晟 陈拥军 30	
基金纵横	
国家自然科学基金联合基金设立与资助管理机制探讨.....朱蔚彤 孟宪平 34	国家自然科学基金委员会对西欧国家组织间合作研究项目的组织、评审与资助.....范英杰 鲁荣凯 38
依托科学基金,开创国际合作新局面.....陈孝政 王丹 李一军 41	进一步贯彻落实《国家自然科学基金条例》若干问题的思考.....王国骞 韩宇 43
中国科学院昆明植物所“十一五”期间国家自然科学基金管理工作综述.....李晓贤 李志坚 46	2001-2011年间内蒙古师范大学国家自然科学基金资助情况统计与分析.....咏梅 金良 董杰 杨宇平 49
国家自然科学基金精细化管理的若干做法与体会.....张力 52	中国科学院上海硅酸盐研究所“十一五”期间国家自然科学基金管理工作回顾.....张明 王东 54
科学基金项目评审专家科研信誉综合评价体系研究:基于“能力-行为-绩效”的视角.....黄进 57	西南大学“十一五”期间国家自然科学基金资助项目增长驱动分析.....张蔚 61