



教育部大学计算机课程改革项目规划教材

丛书主编 卢湘鸿

计算思维应用实例

赵 宏 主编

王恺 高裴裴 张健 李妍 刘爽 编著

清华大学出版社





教育部大学计算机课程改革项目规划教材

| 丛书主编 卢湘鸿 |

计算思维应用实例

赵 宏 主编

王恺 高裴裴 张健 李妍 刘爽 编著

清华大学出版社
北京

内 容 简 介

本书将信息技术与人文科学有机地结合起来,目标是培养大学生主动使用计算机解决问题的意识和计算思维的能力,提高他们对计算可以影响学科发展的认识,同时避免非计算机专业的学生陷入学习一门程序设计语言的苦恼和困惑中。书中给出了多个信息技术在促进人文科学研究和发展的具体方法和实例,涉及文本挖掘、虚拟现实、数字图书馆、计算机辅助心理学、计算机辅助历史研究、计算机辅助翻译、计算机辅助艺术研究、数字人类文化遗产、政府信息化等方面的内容。

本书既可作为高等院校或高职高专“人文类”学科学生的大学计算机基础课程的教材,也适合作为信息技术与人文科学相结合的科普读物。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

计算思维应用实例/赵宏主编;王恺等编著. —北京:清华大学出版社,2015

教育部大学计算机课程改革项目规划教材

ISBN 978-7-302-38583-7

I. ①计… II. ①赵… ②王… III. ①计算方法—思维方法—高等学校—教材 IV. ①O241

中国版本图书馆 CIP 数据核字(2014)第 274379 号

责任编辑:谢琛 李晔

封面设计:常雪影

责任校对:时翠兰

责任印制:刘海龙

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课件下载: <http://www.tup.com.cn>, 010-62795954

印 装 者:保定市中华美凯印刷有限公司

经 销:全国新华书店

开 本:185mm×260mm

印 张:20

字 数:487千字

版 次:2015年3月第1版

印 次:2015年3月第1次印刷

印 数:1~2000

定 价:35.00元

产品编号:055383-01

序

以计算机为核心的信息技术的应用能力已成为衡量一个人文化素质高低的重要标志之一。

大学非计算机专业开设计算机课程的主要目的是掌握计算机应用的能力以及在应用计算机过程中自然形成的包括计算思维意识在内的科学思维意识,以满足社会就业需要、专业需要与创新创业人才培养的需要。

根据《教育部关于全面提高高等教育质量的若干意见》(教高[2012]4号)精神,着力提升大学生信息素养和应用能力,推动计算机在面向应用的过程中培养文科学生的计算思维能力的文科大学计算机课程改革、落实由教育部高等教育司组织制订、教育部高等学校文科计算机基础教学指导委员会编写的高等学校文科类专业《大学计算机教学要求(第6版——2011年版)》(下面简称《教学要求》),在建立大学计算机知识体系结构的基础上,清华大学出版社依据教高司函[2012]188号文件中的部级项目1-3(基于计算思维培养的文科类大学计算机课程研究)、2-14(基于计算思维的人文类大学计算机系列课程及教材建设)、2-17(计算机艺术设计课程与教材创新研究)、2-18(音乐类院校计算机应用专业课程与专业基础课程系列化教材建设)的要求,组织编写、出版了本系列教材。

信息技术与文科类专业的相互结合、交叉、渗透,是现代科学技术发展趋势的重要方面,是新学科的一个不可忽视的生长点。加强文科类专业(包括文史法教类、经济管理类与艺术类)专业的计算机教育、开设具有专业特色的计算机课程是培养能够满足信息化社会对文科人才要求的重要举措,是培养跨学科、复合型、应用型的文科通才的重要环节。

《教学要求》把大文科的计算机教学,按专业门类分为文史法教类(人文类)、经济管理类与艺术类等三个系列。大文科计算机教学知识体系由计算机软硬件基础、办公信息处理、多媒体技术、计算机网络、数据库技术、程序设计、美术与设计类计算机应用以及音乐类计算机应用等8个知识领域组成。知识领域分为若干知识单元,知识单元再分为若干知识点。

大文科各专业对计算机知识点的需求是相对稳定、相对有限的。由属于一个或多个知识领域的知识点构成的课程则是不稳定、相对活跃、难以穷尽的。课程若按教学层次可分为计算机大公共课程(也就是大学计算机公共基础课程)、计算机小公共课程和计算机背景专业课程等三个层次。

第一层次的教学内容是文科各专业学生应知应会的。这些内容可为文科学生在与专业紧密结合的信息技术应用方面进一步深入学习打下基础。这一层次的教学内容是对文科大学生信息素质培养的基本保证,起着基础性与先导性的作用。

第二层次是在第一层次之上,为满足同一系列某些专业共同需要(包括与专业相结合而不是某个专业所特有的)而开设的计算机课程。其教学内容,或者在深度上超过第一层次的

教学内容中的某一相应模块,或者拓展到第一层次中没有涉及的领域。这是满足大文科不同专业对计算机应用需要的课程。这部分教学内容在更大程度上决定了学生在其专业中应用计算机解决问题的能力与水平。

第三层次,也就是使用计算机工具,以计算机软硬件为背景而开设的为某一专业所特有的课程。其教学内容就是专业课。如果没有计算机作为工具支撑,这门课就开不起来。这部分教学内容显示了学校开设特色专业课的能力与水平。

这些课程,除了大学计算机应用基础,还涉及数字媒体、数据库、程序设计以及与文史哲法教类、经济管理类与艺术类相关的许多课程。通过这些课程的开设,是让学生掌握更多的计算机应用能力,在计算机面向应用过程中培养学生的计算思维及更加宽泛的科学思维能力。

清华大学出版社出版的这套教育部部级项目规划教材,就是根据教高司函[2012]188号文件及《教学要求》的基本精神编写而成的。它可以满足当前大文科类专业计算机各层次教学的基本需要。

对教材中的不足或错误,敬请同行和读者批评指正。

卢湘鸿

2014年10月于北京中关村科技园

卢湘鸿 北京语言大学信息科学学院计算机科学与技术系教授,原教育部高等学校文科计算机基础教学指导分委员会副主任、秘书长,现任教育部高等学校文科计算机基础教学指导分委员会顾问、全国高等院校计算机基础教育研究会文科专业委员会常务副主任兼秘书长,30多年来一直从事非计算机专业的计算机教育研究。

前言

2006年3月,时任美国国家科学基金会(NFS)计算机与信息科学与工程部主任的周以真(Jeannette M. Wing)教授首先明确地提出并定义了“计算思维”(Computational Thinking, CT)这一概念:计算思维是运用计算机科学的基础概念进行问题求解、系统设计以及人类行为理解等涵盖计算机科学之广度的一系列思维活动。2011年,图灵奖获得者Richard M. Karp提出了“计算透镜”(Computational Lens)理念,其核心是将计算作为一种通用的思维方式,通过这种广义的计算(涉及信息、执行算法、关注复杂度)来描述各类自然过程和社会过程,从而解决各个学科的问题。

在美国,“计算思维”的提出得到了美国教育界和科学界的广泛支持。在美国国家科学基金会启动了“大学计算教育振兴的途径(CISE Pathways to Revitalized Undergraduate Computing Education, CPATH)”,并投入几千万美元的巨资在美国进行计算教育的改革。2008年,美国国家计算机科学技术教师协会(CSTA)在网上发布了得到美国微软公司支持的“计算思维:一个所有课堂问题解决的工具”(Computational Thinking: A problem solving tool for every classroom)报告,对什么是计算思维进行了总结。由于对“计算思维”所发挥的作用取得了共识,在2009年申报的项目中提出了具体的以计算思维为核心的课程改革。美国国家科学基金会启动了以计算思维为核心的重大基础研究CDI计划,旨在使用计算思维(特别是在该领域产生的新思想、新方法)促进美国自然科学和工程技术领域产生革命性的成果。

在美国国家科学基金会的支持下,多所著名高校已经进行了将计算学科与其他学科大学基础课程相融合,计算思维已融于各类大学本科普通教育课程中。例如,在DePaul大学的自由学习项目中,该大学在艺术与文学,哲学探究,宗教,科学探究,自我、社会与现代世界和了解过去6个领域的几十门课程中,将计算思维以实例的形式明确地教给学生,并且根据计算的7大基本原理,即计算、通信、协作、记忆、自动化、评估和设计,明确地列出了每一门课程能够学习到的计算思维的基本原理。卡耐基·梅隆大学的计算机科学学院也在对其入门课程系列进行大的修订,修订包括:为计算机专业和非计算机专业开设的入门课程要推广计算思维的原理;针对软件的高可靠性加强高可信软件开发及方法的学习;考虑到未来程序主要利用并行计算实现高性能,着力培养学生这方面的能力。

2011年,NSF又启动了CE21(The Computing Education for the 21st Century)计划,目的是提高K-14(中小学生和大学一、二年级)学生及老师的计算思维能力。

计算思维也影响着英国的教育。在英国的爱丁堡大学,人们在一系列的研讨会上探索与计算思维有关的主题。每次研讨会,都有不少专家讨论计算思维对他们学科的影响。研讨会上所涉及的学科已延伸到哲学、物理、生物、医学、建筑、教育等各个不同的领域。另外,

英国计算机学会(British Computer Society, BCS)也组织了欧洲的专家学者对计算思维进行研讨,提出了欧洲的行动纲领。

在我国,计算思维的重要性也已引起了科学家和教育界的高度重视,在“计算思维”理念下,我国也开始了大学计算机改革的进程。从2008年开始,教育部高等学校计算机基础课程教学指导委员会在陈国良院士的带领下,组织了将近20场各种类型的专题研讨,从计算思维的基本概念出发,就哲学层面、科学层面以至于教学层面的表达形式进行了深入的讨论,逐步实现计算思维从哲学的表达体系,向教学表达体系的过渡,以提高国内计算思维领域的科学研究和计算机教育的水平。

2010年7月,北京大学、清华大学、西安交通大学等9所“985工程”高校在西安召开了首届“九校联盟(C9)计算机基础课程研讨会”,发布了《九校联盟(C9)计算机基础教学发展战略联合声明》,正式拉开了我国大学计算机课程改革的序幕。该声明明确了要旗帜鲜明地把“计算思维能力的培养”作为计算机基础教学的核心任务。

2012年8月,教育部高教司设立以计算思维为切入点的大学计算机课程改革项目。该项目通过3项“第一类:大学计算机课程系统性规划研究项目”和19项“第二类:大学计算机系列课程及教材建设项目”的研究和建设,以及进一步提高计算思维在大学计算机基础课程教学中的教学理论水平和实践水平。

2013年4月,北京大学、清华大学、厦门大学等43所院校在厦门召开研讨会,并达成共识——“大学计算机研讨43院校厦门会议共识”。该共识旗帜鲜明地提出了“建设大学计算机体系”、“进一步确立大学计算机基础课程的基础地位”和“在坚持面向应用的过程中培养计算思维”等大学计算机改革方向。

2013年5月,教育部高等学校大学计算机课程教学指导委员会的深圳会议,发布了旨在大力推进以计算思维为切入点的计算机教学改革的宣言——“计算思维教学改革宣言”。宣言明确了“科学思维能力的培养是教育的最重要和最基本的目标之一”,“通过以计算思维为切入点的计算机课程改革,大胆扬弃现有的教学观念和方法,建设适应时代要求的新的教学体系”以及“在这项改革中,我们面临的最大挑战就是构建培养计算思维能力的教学体系”等问题。

2013年7月,在第二届“计算思维与大学计算机课程教学改革研讨会”上,来自全国各高校的360余名教师参加了此次研讨会。会上进行了成果分享,同时,陈国良等众多与会代表在大会上也表达了“计算思维是潜移默化的培养,不应该为计算思维而计算思维”,“计算思维过热,不要再提计算思维了”等观点。

在计算思维理念下,我国大学计算机相关课程的改革取得了一些标志性的和有一定影响力的成果。2011年,陈国良等认为大学第一门计算机基础课程是计算思维培养的一个关键,初步构建了以计算思维为核心的“计算思维导论”课程,并给出了该课程任务、基本要求,教学内容和教学方法。2012年,陈国良、王志强等出版了《计算思维导论》教材,并在深圳大学开设了“计算思维导论”课程。唐培和也出版了《计算思维导论》教材。

2012年,李晓明在全国多所高校开设了“网络、群体与市场”课程,并于2013年开始开设了相应的MOOC课程。该课程从交叉学科的角度出发,综合运用经济学、社会学、计算与信息科学以及应用数学的有关概念与方法,考察网络行为原理及其效应机制。课程讲解了一些社会学和经济学的经典问题实例是如何转变为计算机可以解决的形式,完全脱离了传

统的计算机教材里对系统讲述计算机专业知识的讲授路线。该课程是培养计算思维的一个有益的尝试。

2013年2月,战德臣等通过构建计算思维教育空间——计算之树,从计算技术与计算系统的发展角度阐述了“核心”的计算思维,给出了大学计算机所面对的知识空间,进而通过分析非计算机专业学生未来对计算思维能力的需求,给出了大学计算机课程教学的一个内容体系方案。

2013年4月的厦门会议上,桂林电子科技大学的董荣胜教授做了“计算思维的表述体系(草案)”的报告,以及2013年7月,陈国良、何钦铭等在第二届“计算思维与大学计算机课程教学改革研讨会”上,公布了“计算思维教学改革白皮书(征求意见稿)”。他们构建了具有8类基本计算原理(计算、抽象、自动化、设计、通信、协作、记忆和评估)的计算思维的表述体系,同时也将这8类基本计算原理的关注点及涉及的核心概念进行了梳理。该计算思维表述体系参考了CC1991的12个核心概念和周以真的计算思维的基本概念,借鉴了Denning在“伟大的计算原理”的分类方法。其创新点在于:

(1) 白皮书在Denning的7类计算原理基础上增加了“抽象”,更好地提炼了计算思维的本质;

(2) Denning的7大计算原理是并列关系,陈国良等的计算思维表述体系将8大计算原理划分为3类,更好地描述了各原理的抽象层次和相互关系。

正如李廉教授所言,随着云计算、社会网络、物联网、普适计算、移动通信这些新技术的迅速发展,使得人们去编制一个程序的任务将会被寻找一个程序的任务所替代。对于大多数人所从事的工作而言,理解问题,并在云平台上找到解决问题的工具,其现实意义可能会远远大于自己动手制作解决问题的工具。目前大学计算机改革的一个重要特征是:在非计算机专业的人才培养目标中,如何更好地实现专业化和信息化相融合的模式,提升未来社会对于计算机的理解和应用的整体水平。在这样的时代背景下,究竟给学生讲什么、怎么讲,是摆在广大计算机基础教育者面前的一个尖锐的问题。

总体来说,目前我国在大学计算机课程中培养“计算思维”的改革仍处于一个摸索阶段。在内容和方法上的突破性改革成果还不明显,还没有形成一套或几套科学的、得到普遍认可的、操作性强的大学计算机课程体系,更缺少与各学科紧密结合、将计算基本原理到专业课程相融合的研究成果,真正全方位进行大学计算机课程深化改革的高校也几乎没有。

在2012年教育部启动了大学计算机课程改革项目中,南开大学承担了“基于计算思维的人文类大学计算机系列课程及教材建设”的一个子项建设,建设了人文类大学生的公共计算机基础课程一个包括4门课程的课程体系。这4门课程分别是:“大学计算机基础”、“大学计算机应用教程”、“计算思维应用实例”和“大学计算机应用实践”。

本教材由担任南开大学公共计算机基础教学部的教师结合多年的教学经验和大学计算机课程教学的发展,面向我国高校人文类大学学生,力争在有限课时学习了解和学习一些计算原理和方法,同时具有自觉使用计算思维去解决实际问题的思维方式。赵宏负责第1章和第2章的编写并统编全书,王恺负责第3章的编写,赵宏、刘爽和高裴裴负责第4章的编写、张健负责第5章和第8章的编写(其中第5章涉及的CBDB数据库目前仅有英文和繁体中文版本,因此相应的界面图也是对应的英文和繁体中文界面)、高裴裴负责第6章和第

7 章的编写、李妍负责第 9 章和第 10 章的编写。

本书参考了大量的国内外论文、网络资源和书籍。由于能力和水平的限制,书中出现的不妥乃至错误之处,敬请阅读本书的老师、同学和读者指正。

在本书的编写过程中,得到了清华大学出版社的大力支持。在此表示真诚的感谢!

编 者

2014 年 5 月于南开园

目 录

第 1 章 人文计算概述	1
1.1 人文计算的基本概念	1
1.1.1 人文计算产生的背景	1
1.1.2 人文计算的定义及特点	2
1.1.3 人文计算研究的内容	3
1.1.4 人文计算的发展历史	4
1.2 人文计算研究现状	5
1.2.1 整体情况	5
1.2.2 美国情况	5
1.2.3 欧洲情况	6
1.2.4 日本情况	6
1.2.5 港台情况	6
1.2.6 大陆情况	7
1.3 人文计算的研究应用前沿	8
1.3.1 历史学方面的基于 GIS 的历史地理可视化	8
1.3.2 文学方面的文本挖掘与 TEI 标准	8
1.3.3 语言学方面的基于大型语料库的语料库语言学	9
1.3.4 舞蹈方面的视频捕捉、运动分析与虚拟现实再现	10
1.3.5 考古学方面的图像分析、色彩还原和数字重建	10
1.3.6 数字图书馆、博物馆和网络数据库	11
1.4 数字资源的概念、检索及获取	12
1.4.1 数字资源的概念	12
1.4.2 数字资源的检索	13
1.4.3 主要数字资源	18
参考文献	22
第 2 章 文本挖掘	23
2.1 文本挖掘的基本概念	23
2.1.1 文本挖掘的背景	23
2.1.2 文本挖掘的概念	25

2.1.3	文本挖掘的任务	27
2.1.4	文本挖掘需要研究的课题	29
2.1.5	文本挖掘系统的结构	31
2.1.6	文本挖掘应用概述	33
2.2	文本挖掘关键技术的应用	37
2.2.1	信息抽取技术应用	37
2.2.2	文本分类技术应用	38
2.2.3	文本聚类技术应用	40
2.2.4	关联规则技术应用	41
2.2.5	模式发现与可视化技术应用	41
2.3	文本挖掘应用实例	43
2.3.1	《红楼梦》作者的鉴定研究	43
2.3.2	谷歌流感趋势预测	46
2.4	文本挖掘工具	48
2.4.1	文本挖掘工具	48
2.4.2	典型开源文本挖掘工具比较	50
2.4.3	使用文本挖掘工具的建议	54
2.4.4	Weka 工具及简单应用示例	55
	参考文献	61
第3章	虚拟现实	63
3.1	概述	63
3.1.1	初识虚拟现实	63
3.1.2	虚拟现实的基本概念	64
3.1.3	虚拟现实的技术基础	65
3.1.4	虚拟现实的交互工具	66
3.1.5	虚拟现实的发展历史	70
3.1.6	虚拟现实工具	74
3.2	虚拟现实的应用	75
3.2.1	虚拟现实应用概况	75
3.2.2	虚拟博物馆	80
3.2.3	虚拟现实在医学中的应用	84
3.2.4	虚拟现实在室内设计中的应用	90
3.2.5	虚拟现实在实验教学中的应用	93
3.3	虚拟现实的具体实例	97
3.3.1	相关软件工具简介	97
3.3.2	卧室 3D 建模实例	98
3.3.3	家具 3D 建模实例	99
	参考文献	101

第 4 章 数字图书馆	104
4.1 数字图书馆的基本概念及发展概况	104
4.1.1 数字图书馆的基本概念	104
4.1.2 数字图书馆的产生和演变	105
4.1.3 数字图书馆的发展概况	107
4.1.4 数字图书馆的主要优势	112
4.2 数字图书馆的关键信息技术	112
4.2.1 信息存储技术	112
4.2.2 数据库、网络通信技术	115
4.2.3 信息检索技术	115
4.2.4 虚拟现实技术	117
4.2.5 云计算技术	118
4.3 基于 3G/4G 网络的移动数字图书馆	120
4.3.1 移动电子阅读用户的需求特征	120
4.3.2 国内外移动数字图书馆实践现状	121
4.4 数字图书馆实例	123
4.4.1 中国国家数字图书馆	123
4.4.2 使用中国国家数字图书馆简单示例	128
4.4.3 个人数字图书馆	132
4.4.4 全国主要城市数字图书馆	134
参考文献	141
第 5 章 计算机辅助历史研究	142
5.1 计算机辅助历史研究技术基础	142
5.1.1 信息技术与历史学之间的关系	142
5.1.2 信息技术在历史研究中的应用领域	145
5.1.3 信息技术对历史研究的特殊意义	150
5.2 信息技术辅助历史研究实例——CBDB 数据库	152
5.2.1 数据库技术基本概念	152
5.2.2 CBDB 的概述	153
5.2.3 CBDB 的使用	159
5.2.4 CBDB 的分析工具	167
参考文献	171
第 6 章 计算机辅助心理学研究	172
6.1 计算机辅助心理学研究概述	172
6.1.1 心理学的基本概念及其与计算机科学的关系	172
6.1.2 心理学与计算机技术相结合的研究领域	174

6.2	计算机技术在心理学研究中解决的主要问题	176
6.2.1	心理学实验、测试的操作与控制	176
6.2.2	获取研究数据	179
6.2.3	处理、分析、研究数据	180
6.2.4	心理过程的模拟和模型化	181
6.3	心理学中的眼动行为研究方法及应用实例	182
6.3.1	眼动与心理学	182
6.3.2	眼动跟踪技术与眼动记录仪	183
6.3.3	眼动记录方法在心理学研究中的应用实例	189
6.4	心理学研究中的情感计算及应用实例	197
6.4.1	情感计算与心理学	197
6.4.2	表情与微表情	199
6.4.3	表情识别与人脸运动编码系统(FACS)	200
6.4.4	人脸微表情识别实例	203
	参考文献	207
第7章	计算机辅助翻译	208
7.1	计算机辅助翻译技术概述	208
7.1.1	机器翻译(MT)的概念	208
7.1.2	计算机辅助翻译的概念	211
7.1.3	计算机辅助翻译的性能	214
7.2	计算机辅助翻译的核心技术——翻译记忆	215
7.2.1	翻译记忆的概念和基本功能	215
7.2.2	术语库管理	218
7.2.3	翻译记忆数据库的分离与共享	221
7.3	用搜索引擎整合网络翻译资源	222
7.3.1	网络语言资源	222
7.3.2	网络蜘蛛搜索引擎	224
7.3.3	网络翻译资源的搜集整理	225
7.4	常用计算机翻译工具	226
7.4.1	金山快译(MTS)	227
7.4.2	雅信	232
7.4.3	TRADOS	240
	参考文献	246
第8章	信息技术辅助艺术研究	247
8.1	信息技术对当代艺术的影响	247
8.1.1	艺术传播者范围变广	248
8.1.2	艺术载体的改变及影响	249

8.1.3 艺术研究和传播的受众的变化·····	252
8.2 独特的数字艺术·····	253
8.2.1 信息技术辅助音乐研究·····	253
8.2.2 信息技术辅助舞蹈研究·····	256
8.2.3 计算机辅助绘画研究·····	259
8.2.4 信息技术辅助电影研究·····	261
参考文献·····	267
第9章 人类文化遗产数字化·····	268
9.1 人类文化遗产数字化概念及其必要性·····	268
9.2 人类文化遗产数字化保护现状·····	269
9.3 人类文化遗产数字化技术基础·····	270
9.3.1 立体影像技术·····	270
9.3.2 数字动画技术·····	271
9.3.3 GIS 三维可视化技术·····	272
9.3.4 人机交互技术·····	273
9.4 人类文化遗产数字化实例·····	274
9.4.1 基于立体影像技术的文化遗产实例·····	274
9.4.2 基于数字动画技术的文化遗产实例·····	277
9.4.3 基于 GIS 的文化遗产实例·····	281
参考文献·····	284
第10章 政府信息化·····	286
10.1 政府信息化的含义和意义·····	286
10.1.1 政府信息化的内涵·····	286
10.1.2 政府信息化的意义·····	287
10.2 政府信息化的先进经验·····	289
10.2.1 美国政府信息化建设成功经验·····	290
10.2.2 上海市政府信息化建设·····	292
10.3 政府信息化的主要内容·····	293
10.3.1 完善政府信息资源体系·····	293
10.3.2 科学引导政府信息消费·····	294
10.3.3 深层次开发信息资源·····	295
10.3.4 转变政府职能,提供社会急需的政务信息·····	297
10.3.5 建立面向信息化社会的知识管理体系·····	299
10.4 政府信息化技术基础·····	300
10.4.1 政府信息化建设技术架构·····	300
10.4.2 办公平台技术支撑系统·····	301
参考文献·····	305

第 1 章

人文计算概述



导读

在现代,信息科学的研究成果渗透到了传统的人文科学之中,并为传统人文科学的研究带来了新的研究方法和研究理论,这种改变冲击着传统人文研究的方式和维度,也冲击着传统人文精神的视野和语境,最终使信息技术与人文研究逐渐融合而形成了一个新兴的研究领域——人文计算(Humanities Computing)或“数字人文”(Digital Humanities)。这一新兴研究领域的产生充分体现了计算思维在人文科学中的渗透和应用。

本章主要介绍了“人文计算”的基本概念、发展历史、研究现状和研究趋势。目的是使读者认识到,“人文计算”在为传统人文学科研究提供了全新的方法和理论的同时,也对人文学科的研究提出了更高的要求。最后还提供了利用网络获取数字资源的主要途径和方法,以提高人文科学科研人员和大学生检索信息的能力;并列出了主要的中文数字资源和外文数字资源,以满足他们在人文科学的学习研究中对数字资源的需求。

1.1 人文计算的基本概念

1.1.1 人文计算产生的背景

科学是人类对客观世界认识的结晶,学科是对科学的细分,是人类认识客观世界实践经验的概括和总结。科学的基本使命是认识客观世界,是以概念、范畴、原理、定律等理论形态,按照周围世界的本来面貌反映周围世界,正确地揭示客观世界所有现象和过程的本质的、必然的规律。为了研究上的方便,人们按不同的标准将这些概念、范畴、原理、定律等分为不同的类别,构成了今天的学科。

学科之间相互交叉渗透是当代科学发展的一个重要趋势,学科交叉的形成与发展极大地推动了经济与社会的发展。从历史上看,学科交叉能产生新的学科生长点,是获得原创性科学成果的重要途径,也是解决重大技术、社会问题的必然选择。

人文计算是一个将现代计算机和网络技术深入应用于传统的人文研究与教学的新型跨学科研究领域,它的产生与发展得益于数字技术的进步及其在科学领域的普及应用。20世纪60年代以来,在计算机媒介的支撑下,人文知识的获取、分析、集成和展示都在发生重大变化。目前,已有海量的图书、报纸、期刊、照片、绘本、乐曲、视频等人文资料被数字化,并在网络上被提供给大众获取和使用。面对这种日益强化的数字化趋势,人文学者急需相应的工具和平台来对这些数字化人文资料进行组织、标引、检索和利用,以保证人文研究的持续

性、一致性和高效性。人文计算研究的正是这些数字媒介被应用于人文社科领域的过程,以及它们给人文科学和人文知识带来的变化和影响。

1.1.2 人文计算的定义及特点

关于什么是“人文计算”,目前还没有一个权威且统一的定义。

美国伊利诺伊香槟分校图书情报学院院长约翰·昂斯沃斯(John M. Unsworth)教授(见图 1.1)是国际人文计算领域的知名专家,曾与他人合著人文计算入门经典图书 *A Companion to Digital Humanities*。他在 *WHAT IS HUMANITIES COMPUTING AND WHAT IS NOT?* 一文中对“人文计算”的概念和范围进行了界定。他认为,“‘人文计算’是一种代表性的实践、一种建模/模拟的方式,一种推理、一个本体论约定。这种代表性的实践可分为两个方面:一方面是高效的计算;另一方面是人文沟通”。



图 1.1 John M. Unsworth

人文计算的特点可以概括为以下几个方面:

(1) 是一个跨专业的学科交叉领域。

人文计算的研究项目和研究团队常常既包括传统人文领域(哲学、历史学、文学、语言学、艺术学、人类学等)的研究者,还包括精通计算机技术和多媒体技术的专家学者。在这两类人员的协作下,诸如数字仓储、文本挖掘、多媒体出版、数字图书馆、信息可视化、虚拟现实、地理信息系统等多种信息技术开始在人文领域得到深入应用。正如 Unsworth 教授所说,“一方面是高效的计算,另一方面是人文沟通”,人文计算研究就是将现代计算机技术、电子技术、网络技术等应用到文学、历史学、艺术学、考古学等传统人文学科中,这种应用在几十年前的人文学科研究中几乎是无法想象的,以至于到今天仍然有着持续不断关于“经典的阅读”与“阅读的经典”之争。但不可否认,这种应用为传统人文学科研究提供了全新的方法和理论。

(2) 研究的对象是可计算的基础数字化对象。

人文计算研究的核心在于将研究对象数字化以支撑、保障和创新人文科学研究的内容、方法和模式,研究对象包括各种可计算的基础数据对象,如自由文本、格式化数据、图像、声音等。针对这些数据进行的计算包括文本分析与检索、地图可视化、音视频检索等。人文计算的方法随着信息技术的发展而不断变化,因此其研究边界相对不固定。

(3) 关注的是信息技术对于人文研究的本体实践意义。

人文计算研究最重要的实践是将信息技术方法借鉴到人文研究中。在数字化和技术化的社会,人们逐渐开始质疑一些传统的人文科学研究方法,因其过于注重思想、理论的一致性而忽视研究中的科学性和实践性。美国心理学家斯金纳在其著作中写到:“亚里士多德丝毫不懂现代物理学和生物学,但苏格拉底和他的朋友们却能不大困难地理解今天关于人的问题的大部分讨论。就技术而言,我们在控制物理世界与生物世界上已取得了巨大的进步,但在政治、教育以及许多经济领域方面却并无多大进展(虽然这类技术所要适应的社会环境与古代已大不相同)。”虽然斯金纳对于传统人文学科研究的批评过于片面和尖锐,但不可否认的是,他对于传统人文学科研究的认识是十分深刻的,他在文中所提出的观点在世界

范围内的人文学研究领域都引起了不小的震动。

(4) 对人文学科研究提出了更高的要求。

在科研人员的努力下,诸如数字仓储、文本挖掘、多媒体出版、数字图书馆、信息可视化、虚拟现实、地理信息系统等多种信息技术已经开始在人文领域得到一定程度的应用,并初步展示出其让人惊叹的成果。人文数字化的同时,也对人文科学的研究提出了更高的要求。对于人文科学的专家、学者,特别是未来要从事人文科学研究的大学生,必须在这些方面进行积极的学习和探索,未来的人文科学研究人员必将是对技术知识和人文知识都有较高要求的跨多学科的人才。

人文计算领域的先驱,意大利著名人文学者 Roberto Busa 认为人文计算化的最重要结果并不是加速传统人文研究的速度,而是给传统的人文研究提供新的研究方法和研究范式。这一解释表明人文计算的产生在本质上属于一种方法论和研究范式上的创新,其目标是将现代信息技术融入于传统的人文研究与教学过程中,从而在根本上改变人文知识的获取、标注、比较、取样、阐释与表现方式。

1.1.3 人文计算研究的内容

为了明确人文计算研究的路线和研究内容,国际文学与语言学计算联合会(Association for Literary and Linguistic Computing)主席 Antonio Zampolli 教授曾于 2001 年 4 月召集多个学科的研究者开展了一次综合性的人文计算研讨会,随后发布了 Pisa 报告。在此报告中,David Robey 教授绘制并发表了一幅有关“人文计算”的全景知识地图(<http://www.allc.org>),如图 1.2 所示。

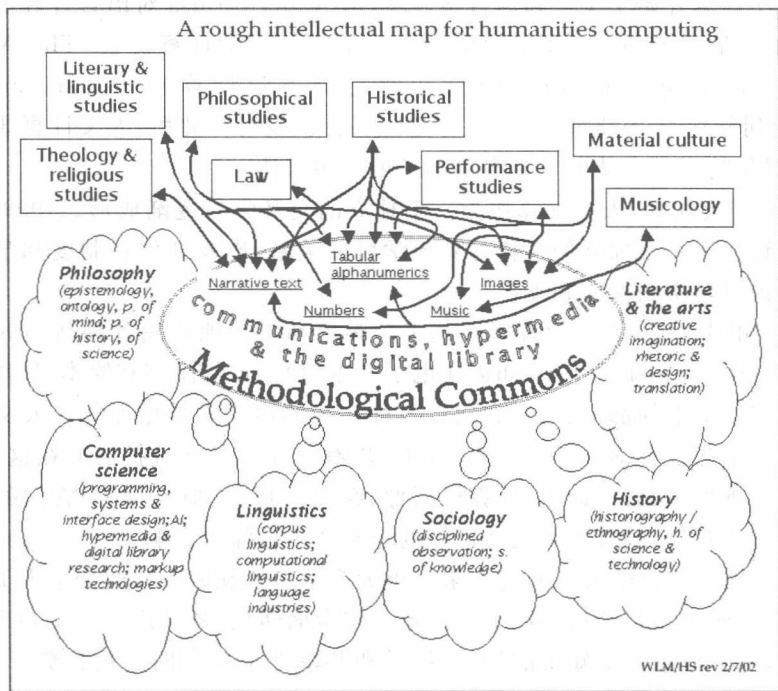


图 1.2 人文计算全景知识地图