

# 用大数据技术促文化资源管理

文·姜念云 本报记者 滕继濮

文化资源是体现一个国家文化实力的核心要素,也是国家文化及文化产业发展的基础和源头。我国对各类物质和非物质文化资源数字化工作的开展,为我们利用大数据分析等先进技术,加强对中华文化的充分认知和深入挖掘利用提供了前所未有的契机和条件。

如何利用大数据等先进信息技术的手段,进一步加深对中华民族几千年文明史的系统了解和认知,在更好地支撑我国文化产业发展的同时,促进中华文化血脉的传承延续,是科技部高技术中心在组织开展科

技部委托的“国家文化科技创新战略研究”项目过程中意识到的重大问题。

日前,由该中心主办的主题为“大数据技术在文化资源管理中的应用”的“西苑沙龙”在京召开。沙龙邀请了国内研究大数据技术和模式识别、中文信息处理、知识挖掘等先进信息技术以及有关文化艺术创作、文化企业等方面的专家学者,围绕着大数据技术在加强文化资源管理中的作用、需求和技术途径等方面问题,展开了一场跨学科、跨领域的深刻讨论。

## 历史文化资源大数据与分析技术

大数据是指数据量大,但究竟怎样的量才算大,目前并没有统一的定义。一般认为,大数据的数量级至少应该达“太字节”(TB)以上。因为达到了这个量级以上的数据,利用现有IT技术和硬件工具将难以在可容忍的时间内,对其进行有效感知、获取、管理、处理和利用,必须要开发新的数据管理和处理软硬件技术,才能满足应用需求。

随着各类数字化文化资源信息的不断产生,各类数字化文化资源库的不断建立与完善,在客观上为我们建立了一个庞大的、具有大数据特征的数据库和资源库。这为我们进一步利用大数据分析等先进的信息技术手段,实现对这些文化资源信息的整合、梳理、分析、凝练,提供了前所未有的基础和条件。

其中,基于数字化的历史文化资源大数据是计划地对各类历史文化资源数字化所形成的大数据信息。对这类数据的有效管理和充分挖掘、利用,或许

是大数据及其分析技术最为重要的应用角度和需求。

1947年春天,一名牧童在死海附近的一个山洞里发现了一些残缺不堪的手稿。1948年至1956年间,研究人员又在死海西北基伯昆兰旷野陆续发现11个藏有手稿的洞穴。十多年间,考古学家发掘出近4万古卷残片,种类多达600多种。这些手稿大部分用希伯来文写在羊皮上,公元前二三世纪到公元70年间写成,记录了大量关于《圣经》和基督教起源的重要信息。这在近代考古史上非常罕见,被称为20世纪最伟大的考古发现之一。这就是著名的“死海古卷”。

利用数据处理技术,“死海古卷”机器近几年自动修复的效率,已经与数百人类专家过去一个世纪的成果相当。另外,梵高、勃鲁盖尔等大画师作鉴别精准度达到了95%以上。这些都是目前利用大数据分析技术进行画作鉴别、古文修复、历史文物分析等取得的惊人成果。

## 中华文化与信息孤岛

一个国家、民族的文化资源是其文明发展的历史过程中沉积形成的独有资产,具有唯一性和不可扩展等特点。因此,具有不可估量的文化、经济和社会价值和意义,是代表一个国家文化软实力的核心内容和象征要素,也是各类文化艺术产品创作的基础资料和源泉。

青花瓷背后还有哪些我们不知晓的特点?手工织造对于现代工艺是否有所启发?经过几千年的文化沉淀,先辈给我们留下了无数的财富,但仅靠少数的专家一件一件文物的深究,对于进一步发掘文明成果来说,是否仍不成体系,不甚具体,不够全面?或者在这些文化瑰宝里,有着更多我们不理解的东西?

大数据分析技术在历史文化分析研究的成果,对于我们进一步加深对中华民族文明发展的历史,认知、辨识中华文化“基因”,延续文脉,明确我国文化建设的重点内容,制定国家文化发展战略具有重要的意义。

第三次文物普查表明,现在在不可移动文物总数已经

达到了76.6722万件;在可移动文物,预计在3亿件组的规模。每年我国有100个博物馆新增,平均三天就出现一个博物馆。目前很多博物馆、图书馆,以及非物质文化遗产保护组织与部门,正在以不同方式、为不同的应用目的,开展着对各类历史文化资源的数字化工作,客观上形成了一个前所未有的难得的中华文化大数据汇集。

但由于这些数字化资源分散在不同的单位、部门,且基本没有统一的格式标准,形成了一系列新的“信息孤岛”。出于各种原因,各自独守一摊,各类数据如果无法汇集到一起,也就难以充分发挥其应有的作用。

文化部民族民间文艺发展中心李松认为,传统文化资源数字化整合技术应用能力低下,体制机制制约使得资源碎片化,行政制约非常严重。国家文物局博物馆与社会文物司副司长罗静表示,在数据融合方面,信息的“孤岛”也是非常严重的。就海量数据来说,如果没有通过一定的融合、整合,其实它还是海量数据,而不是大数据。此外,文化数据整体框架上面也存在碎片化的状态。

## 文化与科技的共同任务

如何消除这些信息孤岛?如何有效整合各自整理出的数字化文化资源?如何利用大数据分析技术对这些资源的价值进行进一步的挖掘和利用?

在现行体制下,专家期望通过一定的技术手段,以及适宜的共享共建机制,构建一个实际或虚拟的数据交汇中心或平台,整合各类数据资源。以便在此基础上,进一步发挥计算机中文信息处理、模式识别、知识挖掘等大数据分析技术的优势,加强文化研究,促进我们对中华文化的认知和保护利用。

也正因此,面向各类文化研究、文化艺术创作、文化管理等用户提供更为优质、高效的信息服务,便成为了一个需要文化与科技相关领域共同探讨、推进的任务。

专家们认为,根据历史文化资源数据信息来源及其结构多元、主要应用领域、方式与用户多样等特点,面向历史文化资源管理的大数据分析系统,是数据来源和应用端开放的,能够实现数据提供相关主体和各类用户共建共享的数据管理平台。

为建立开放共享的数据交汇中心,需要建立技术系统框架,并需解决的四面体的技术问题:建立数字化标准;设计统一的跨平台中间件;实现高效的数据共享;加强对应用于文化资源大数据分析的基础设施的集成利用。

同时,加强各类数字化文化资源的共享利用,不但需要有适宜的技术支撑,同时也需要面对应用需求和客观现实条件,加强顶层设计,在组织管理模式和保障措

施上有所创新。

首先应加强对我国文化数字资源管理的系统规划。明确总体目标、任务和发展战略,提出有利于促进国家文化资源信息大数据管理利用的,由国家和社会相关机构共享共建的组织机制、商业模式和技术框架与标准。

第二可通过国家层面的文化资源管理与共享工程的组织,引导有关图书、文物、非遗保护部门以及各类文化企业、研究机构等有关单位和组织,共同围绕国家文化资源管理的发展战略和目标,开展各类历史文化数字资源库的建设,及其价值挖掘研究和综合利用产品开发。

第三可通过建立国家级的文化资源整合应用虚拟平台等方式,从技术上为各类、各源的文化资源大数据的整合、交汇和综合利用,以及安全维护、利益分配和知识产权保护等提供支撑。

第四是开展形式多样的应用示范,选择若干重点文化主题,组织开展系统深入的文化研究;并充分发挥市场和各类文化企业的作用,开展基于中国历史文化资源的形式多样的文化、艺术服务产品开发,为提升我国文化产品的文化内蕴和中华文化的国际传播力和影响力提供支撑。

“文化原动力是国家建设的基础性建设,需要科技部跟文化部共同有一些软研究在顶层设计上,然后在架构好的时候给科技一些力量,余下的由业界来做。”中科院自动化研究所研究员徐波说。

(稿件素材由西苑沙龙提供)



从明年开始,谷歌就已经与以色列博物馆合作,先后将5份完整版的死海古卷卷轴扫描图片和5000张图片放上互联网。据悉,这些古卷轴在之前全世界只有5名管理人员被允许接触。

### 专家视点

李松 文化部民族民间文艺发展中心副主任:

我强烈呼吁关注改善文化资源准备的技术环境,包括各类资源抢救、修复、复原、记录的相关技术,提高各类形式媒介记录效能的相关技术,加强大数据管理、清晰、分析的相关技术,进一步融合到各类资源信息的标准化体系建设。

我建议在国家战略层面应该构建大型的文化系统工程,建设文化研究和文化传承的数字化广泛应用体系。

罗静 国家文物局博物馆与社会文物司副司长:

从国外来看,行业的数字化相关工作可以分为几类,政府主导、博物馆主导、NGO主导、企业主导的相关工作。从需求提出、数据采集、数据融合、数据挖掘、智慧应用方面来看,大数据技术在文化资源管理中依旧存在很多问题,整体的框架不清楚,而且大多数文化机构在做的工作都是局部的探索性的工作。

孙一钢 中国国家图书馆研究员:

现在传统图书馆和数字图书馆并行的阶段。数字图书馆很重要,在承担一个国家文化传承的职责。从数字图书馆服务体系来讲,大数据现在面临着非常严重的知识组织理论和系统的研究问题。

徐波 中国科学院自动化研究所研究员:

不管是小数据还是大数据,还是需要有很好的复杂算法来刻画丰富的内在信息。大数据模式识别技术聚焦,一个是怎么感知深度学习,另一个是认知层面的知识图谱和推理,这是两个很共性的问题。

周志华 南京大学计算机软件新技术国家重点实验室教授:

数据挖掘也就是机器学习提供数据分析技术再加上数据库提供数据管理技术一起来成为数据挖掘,这在医学、考古等方面有诸多的应用。关于大数据,有3个误区:有大数据,“模型”不重要;大数据,只要“相关”,不要“因果”;大数据,需搭配新的方法。

金海 华中科技大学计算机技术学院教授:

在海量的知识管理和推理方面,通过海量数据甄

别推演新的文化知识,这些技术为文化创新提供多元文化要素转型升级。云计算是大数据技术的支撑环节,对大数据资源挖掘提供了基础的设施,同时包括大规模的高线数据处理、流式数据处理、内存计算技术等为大数据挖掘提供技术支撑。云计算的主要目的为文化创新提供软硬件的支撑平台。

曾成钢 清华大学美术学院教授:

大数据虽然跟美术没有直接关联,但提供了很多支撑技术,对业务来说至关重要,开辟了新的方法和途径。

关键 上海方正数字出版技术有限公司副总经理:

方正公司利用大数据技术,做了数据收集和聚合的引擎,汇集很多文化教育信息,基于传统的数据库技术,做图书出版版的决策和分析,包括从内容汇聚到最后内容重新组织展现的完整解决方案。

张加万 天津大学国家示范性软件学院教授:

大数据技术在文物保护方面相关的工作现在是在做得较多的,例如文物的三维数字化及应用等。总的来说,有四个方面的工作是非常重大的。第一,如何把馆藏文物的资源和网络资源进行整合,通过建立分布式、异构、虚拟文物知识库及应用平台。不仅要把分析放进去,还要把知识放进去。第二,关于基于大数据准入的标识和鉴别。第三,基于大数据分析的文物劣化机理研究。最后,基于大数据分析的观众和游客行为及个性化服务技术。从人、物、环境的角度进行研究,并且来解决服务质量的问题。

许鲁 中国科学院计算技术研究所研究员:

从文化资源的管理角度,我们需要公平自由的文化共享,需要大数据技术的支撑,存储技术的作用明显。

### 数字

#### 4.28亿人次

全国假日旅游部际协调会议办公室8日发布《2013年国庆假期假日旅游统计报告》显示,国庆假期期间,全国共接待游客4.28亿人次,比2012年中秋国庆节假期增长0.7%(按可比口径,同比增长15.1%);实现旅游收入2233亿元,比2012年中秋国庆节假期增长6.1%(按可比口径,同比增长21.2%);游客人均花费支出521元。

在所实现的2233亿元旅游收入中,民航客运收入82.5亿元,铁路客运收入30.8亿元;39个重点旅游城市实现旅游收入796亿元;其他旅游城市和景区实现旅游收入1324亿元。假日期间,纳入全国假日旅游统计预报体系的北京等39个重点旅游城市接待游客1.33亿人次,其中过夜游客为2569万人次,一日游游客为1.08亿人次。

#### 2000亿元

财政部8日发布的月度彩票销售数据显示,今年1月至8月累计,全国共销售彩票1986.92亿元,比上年同期增加297.26亿元,增长17.6%。其中,8月份单月销售彩票246.18亿元,同比增长24.9%。

财政部表示,8月福利彩票和体育彩票销售量均呈现上升态势,主要是由于部分省市调整和上市快速开奖游戏、竞猜型彩票营销推广力度加大,中福在线部分游戏扩大销售试点范围等原因,乐透数字型、竞猜型和视频型彩票销售量与上年同期相比出现较快增长,使得彩票整体销售量有所增长。

分地区看,8月份,全国各个省的彩票销售量均较上年同期有所增长,其中广东、山东、河北、江西和湖北增加额较多。1月至8月累计,除江苏外,全国其他省份彩票销售量均出现增长。

#### 800件

中央纪委监察部网站8日消息,中央纪委监察部网站开通以来,举报网站统计的网络举报数量呈现明显上升之势。9月2日至10月2日,中央纪委监察部举报网站统计的网络举报数量达2.48万多件,平均每天超过800件,而在今年4月至8月每天只有300件。

网站设立的网民互动专栏也受到极大欢迎,截至10月2日,互动板块已有注册用户4217个,发布留言3500多条。

9月2日,由中共中央纪律检查委员会、中华人民共和国监察部主办的综合性政务门户网站中央纪委监察部网站正式开通。

#### 6458家

山东省东营市现有派出所46个,辖“九小场所”(指小商场、小学校/幼儿园、小医院、小餐饮场所、小旅馆、小歌舞娱乐场所、小网吧、小美容洗浴场所、小生产加工企业。)6458家。市消防支队指导公安派出所实行“五户联防”机制开展“九小场所”消防管理工作。目前,全市已创建8条“九小场所”消防示范街。

“九小场所”每五户结成对子成立一个消防安全联防工作领导小组,负责五家经营户的消防安全工作,开展消防安全互查互督,轮流值班开展消防安全提示和检查。同时,明确各经营户的法人为本单位的消防安全责任人,对本经营单元的消防安全工作全面负责,将消防安全与本单位的日常经营进行同安排、同落实。

## 介电电泳技术首次用于破解“十面霾伏”

科技日报讯(记者林莉君)10月10日,

内蒙古天一集团所属派环境科技有限公司在京召开技术发布会,公布了一系列基于介电电泳(DEP)放大应用技术、具有自主知识产权的PM2.5处理技术。这是DEP在世界上首次应用于微米、纳米粒子的捕获、富集。

利用该技术开发的汽车尾气催化/颗粒物捕获装置,可以使尾气PM2.5的排放浓度减少80%以上,如在怠速状态下尾气平均值

由每立方米214微克减少到每立方米41微克,发动机转速达到2500转时PM2.5瞬间数值由每立方米1195微克减少到每立方米213微克。在工业除尘方面,DEP对PMS以下颗粒物的去除率可以达到93%以上,而能耗方面只有现有处理装置的40%,远低于现有的袋式除尘器及高压静电除尘器。

据介绍,天一集团下属的内蒙古介电电泳应用技术研究院,是世界上唯一成功实现DEP

(介电电泳)产业化技术的研究机构。该机构以介电电泳理论的放大应用为研发方向,将此技术应用到无污染物分离、水处理领域、生物领域、新材料及新能源等领域。此前,世界首例介电电泳技术产业化应用装置在天津滨海新区落户,该设备能将一些黄金尾矿中的贵金属——金分离出来,且连续作业、无污染、零排放,与目前世界同处理量金矿分离装置相比,能耗仅为同类装置的60%,提取率接近100%。

科技日报讯日前,随着满载旅客的D6503次列车缓缓驶出了南昌西站,由中铁四局一公司参建的向(塘)莆(田)铁路正式开通运营。

向莆铁路以客运为主兼顺货运,西起江西省南昌市,东至福建省中部沿海地区的莆田市,全长603.6公里,是赣闽两省联系的重要通道,也是中国中西部地区通往福建省的快速铁路通道。由中铁四局一公司

## 向莆铁路通车完善华东地区铁路网

科技日报讯日前,随着满载旅客的D6503次列车缓缓驶出了南昌西站,由中铁四局一公司参建的向(塘)莆(田)铁路正式开通运营。

向莆铁路以客运为主兼顺货运,西起江西省南昌市,东至福建省中部沿海地区的莆田市,全长603.6公里,是赣闽两省联系的重要通道,也是中国中西部地区通往福建省的快速铁路通道。由中铁四局一公司

承建的向莆铁路JX-2B标段,管段全长24.56公里。含南昌西站、三江镇车站及南昌西动车运用所,合同造价约15亿元,管内有大桥8座,约14.5公里。

中铁四局一公司克服了地方关系复杂、资金紧缺、地质条件复杂等困难,科学安排施工进度,合理调配机械、人力、物资等施工资源,先后按期完成了南昌三江镇站改,三江镇特大桥桥墩地质和南昌西动车所房建工

程等重点、难点项目和控制性工程,为向莆铁路的全线如期开通做出了积极贡献。

日前,向莆铁路已正式通车,它的建成通车将有利于加强和完善华东地区铁路网,加快海峡西岸经济建设,构筑对内联结通道,加强沿海与内陆之间的经济协作和优势互补,并且能够极大改善沿线地区的交通条件和加快老区经济社会发展及沿线旅游资源开发。(文良敏 舒郁仁)

## 易车网购车狂欢季开启

科技日报讯(记者马爱平)继9月9日第一期的“易车惠·购车狂欢季”之后,易车网在9月30日全面引爆第二季购车风潮。据悉,本次活动易车网为广大车友提供了低价格、多重优惠和大奖来回馈广大车友,回馈购车用户的总计金额达到了3500万元。“易车惠·购车狂欢季”是易车网最新推出的线上购车产品,活动携手包括华晨宝马、斯巴鲁、东风日产、广汽三菱、一汽、上海大众、雪佛兰等在内的40余家车型。在活动期间,在易车惠平台下订单,即有机会获得500元的加油卡。此外,在12月19日,在易车惠买车成功的车主,

即可获得赢取奖品的机会,奖品为总价值均在百万以上的携程旅行卡、途家网惠住卡及加油卡。同时,易车网官方微博发布了一套以“脱光要趁早”为主题的平面摄影作品,举办了为期一周的有奖微博活动,每天将有两名幸运网友,赢得1000元的购车基金。