

编者按:大数据是继云计算、物联网之后 IT 界的又一次颠覆性的技术革命,给各行各业的发展模式和决策带来前所未有的革新。在教育领域大数据也正在重塑业务流程、IT 基础设施以及我们对于教育信息、教学资源的捕获与使用方式。面对大数据时代的到来,教育信息化面临哪些机遇与挑战?又将如何应对?本专题从大数据创新智慧校园服务、大数据与智慧教育、大数据时代的区域教育信息化发展以及大数据对教学模式和教学环境的变革等几个方面进行了论述和展望。

大数据创新智慧校园服务

宓 咏,赵泽宇

(复旦大学 校园信息化办公室,上海 200433)

摘要:大数据时代的到来为陷入瓶颈的信息化校园数据利用提供了新的出路。大数据将全面、多源的数据作为一种优化服务的手段纳入视野,对高校信息化服务的发展将产生巨大的影响。文章介绍了复旦大学在智慧校园建设过程中充分应用大数据开展的一些特色工作。

关键词:大数据;智慧校园

中图分类号:TP3

文献标志码:A

文章编号:1673-8454(2013)24-0003-05

一、信息化进入智慧校园时代

智慧校园是高校信息化建设的新阶段,其特征包括移动互联网、物联网在内的网络高速泛在、无缝互联与环境感知、智能终端广泛普及与应用、开放协作的学习科研和工作环境、知识的分享和创新等。智慧校园与互联网新技术结合紧密,利用云计算、虚拟化和物联网等来改变校园用户的交互协作、知识传递和资源共享形式,将学校的教学、科研、管理与资源、应用进行整合,以提高服务的灵活性、全面性和易用性,从而实现校园智慧化服务和管理的新模式。

智慧校园的内涵和最终目的是服务。智慧校园以数字校园为基础,追求资源与应用的高度整合,强调为用户提供个性化的服务,更多地体现服务的智能化和人性化。复旦大学在“十二五”规划中将智慧校园的建设分为如下几个基本要素:

- 1)感知层:对学校的人员、设备、资源的全面感知;
- 2)网络层:各类网络的互联互通,所有感知信息的实时传递;
- 3)数据层:数据的全面集成和智能分析;
- 4)应用层:智能决策、按需服务、灵活应对。

在规划设计中描绘了一个学生未来一天的学习、生活场景,体现了信息化无处不在、服务优质便利的智慧校园个性化体验:

7:00,寝室,手机,来自日历系统的提醒,今日课程:中国历史,8:00-9:40,3教205;办公自动化,13:30-16:00,原定计算中心201,因特殊情况更改为计算中心306。

8:00,教室,平板电脑,无线网络:云存储系统下载本节课讲义与参考资料。

9:50,教室,平板电脑,无线网络:地理信息系统,查询3、4节空闲教室准备自修。

11:40,室外,手机:地理信息系统,懒得排长队,查询人流量较小的食堂。

13:20,路上,手机:通知里提醒了更换教室,还好不至于走错。

13:30,多媒体教室,瘦客户端:VDI云桌面,通过学号登录,系统就是老师预先定制的模板,各种相关程序和素材均准备好了,完成课堂作业后直接通过挂载的云存储系统上交。

15:30,运动场:凭校园一卡通在pos机上验证身份、借用器具,参与体育锻炼。

17:00,食堂:手机通过“i复旦”查看各食堂用餐人数情况和菜单,寻找合适的餐厅。

18:00,路上:通过手机看到图书馆阅览室已经满员了;查看教学楼未安排上课的空闲教室、进入教室自修,打开笔记本,通过校园无线网登录网络课堂参与讨论。

22:00, 寝室, 笔记本电脑: 白天课程的材料不需要U盘, 云存储系统已经同步到本地, 助教批改好的作业也已经包含其中, 课程讨论区里, 老师推荐的一部资料片有热心的同学分享, 直接转存到个人空间等明天看。

在智慧校园的场景设计中, 校园网, 尤其无线网、物联网是承载服务感知的基础设施, 师生用户拥有云端的存储和计算资源, 用户通过手机、平板电脑和智能终端体验服务, 基于知识的资源共享成为学习活动的重要组成部分, 个人服务系统的用户体验决定了服务的接受程度, 信息推送、个人日历、提醒服务等信息服务将彻底整合……归根到底, 智慧校园将极大地改变师生的体验。

复旦大学计划将建设重心集中在这几个方面: 完善无线网络覆盖, 建立以RFID、无线AP、校园卡等为基础的物联网, 研究知识管理与分享为基础的在线教学和科研服务体系, 围绕用户体验提升建立全校多种资源的整合、预约、管理等, 建立服务的能力预测和结果评估机制, 统一学校服务接口, 形成以“i 复旦”移动智能终端应用和个性化“一站式”服务平台为依托的信息化服务门户。

对数据的充分利用是服务优化的有效手段。在数字校园建设期间, 许多高校都开展了数据中心建设, 积累了大量教学、科研、运行管理等方面的数据, 传统的以流程为主体的业务模式所沉淀下来的数据成为新的宝贵资源。如何在数据中发现有价值的信息, 转化为服务优化的依据, 是校园信息化实施过程中所面临的问题。大数据作为解决问题的创新思路, 成为业界的火热话题。在高校信息化面临提升的阶段, 它被引入智慧校园建设, 引导着教育信息化工作者从繁琐的业务中抽身出来, 推动业务服务向更深层次的数据服务模式转变。

二、数据应用的现状与困境

数据是信息化建设的核心资源, 是信息化服务得以推行的基础。很多学校在信息化建设过程中, 虽然都考虑了数据的积累和管理, 但在如何分析挖掘、充分利用方面还不够重视, 很多业务系统早期设计中基本以完成业务为

主, 很少去考虑数据如何利用。以复旦大学为例, 自2001年以来, 建成支持全校师生统一身份认证、共享数据库和URP应用系统的信息化校园环境。据不完全统计, 截至2012年, 学校各业务系统已有结构化数据表近万个, 记录数2亿余条, 核心业务数据规模近400GB; 在非结构化数据方面, 也积累了大量的课件、视频材料, 但这些数据和资源还没有得到深入地利用。这几年随着存储价格的下降, 存储规模在不断提升, 我们已逐步扩容到有近PB的数据存储设备, 通过对个人查询、院系/部门统计和校级决策支持等三个层面数据中心的统一建设规划, 将陆续服务于校园网络公开课、个人云存储、数据备份和分析等信息服务。

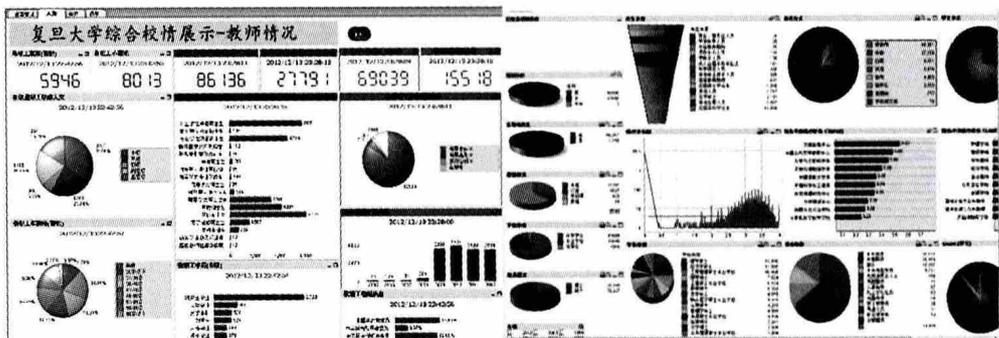
在建设中积累的庞大数据量, 使得通过数据分析获得知识和服务的能力进一步提升, 为全校师生及决策层提供不同的数据服务的迫切性也逐步显现出来。复旦大学积极开展信息化数据应用, 主要提供如下几个方面的服务:

1. 数据整合支持“一站式”服务

信息化“一站式”服务的设计往往会涉及多个部门的多项业务, 为保障服务中心各项业务的顺利流转, 需要多部门的业务数据协同支撑。数据中心将运行过程中逐步积累的业务数据合并、抽取和同步置入综合服务数据库, 通过数据筛选, 面向跨部门业务和个人综合业务形成完整、独立的主题数据库, 提供数据封装和应用接口。在业务流过程中, 无需变更各部门的工作流程, 上一步骤的结果数据利用数据集成技术交换下一步骤的输入数据, 实现信息和流程的无缝对接。基于高度集中的业务数据交换, 建立了包括迎新、离校等跨部门业务的“一站式”服务平台, 以及依据师生角色生命周期形成教务、科研、认识、IT和其他等5个类别70多项业务服务的门户。

2. 集中展示综合数据服务业务部门

综合数据展示服务向职能部门提供自定义或预设主题的查询和展示服务。依托第三方数据展示工具, 根



据各职能部处的数据需求,逐步建设涵盖教职工与本专科生、研究生的6大类17小类的主题数据展示系统,主要包括人员信息类(教职工综合信息统计、学生综合信息统计、本专科生基本信息查询、教职工基本信息查询、研究生基本信息查询),科研类(文科科研数据分析、教师学术表现),教学类(研究生成绩分析、招生数据分析、本科生生命周期数据分析),综合服务类(一卡通分析、图书馆客流分析、综合校情展示、其他信息),资产类(宿舍数据统计、宿舍使用情况统计、资产设备统计),报表相关类(教育部高基报表)等主题。

3. 数据统计和分析辅助校务改革

根据业务部门的要求,对全校的各类数据进行综合分析,通过纵向对比(与本校历来的数据相比较)和横向对比(与其它同类院校的数据相比),发现问题并采取相应处理措施,辅助学校的管理决策,提升学校管理水平。以招生管理为例,复旦大学在过去几年中,信息化部门一直在配合招办开展以数据研究支持招生就业对比分析、指导招生改革的工作。分析过程主要关注国内不同生源地(东部/西部/城镇/农村)入学后的绩点、不同类型生源(自主招生/保送生/委培生/预科生/留学生/港澳台学生)入学成绩和毕业成绩对比、毕业去向对比等,对学生的入学、学习和毕业过程的跟踪分析反馈,为招生改革提供数据支持。

在开展的数据分析和应用工作过程中,一些问题暴露出来。

一是数据的规模不足、形式单一

尽管信息化校园产生了大量的数据,但从实际展示和分析应用中,绝大多数使用的都是结构化业务数据,非结构化数据的应用还未真正得到重视。非结构化数据有各种文档、课程资料、科研数据、创作素材、网络日志、用户发表的网络言论等,往往在校园网和系统运行、师生利用校园网进行教学科研过程中产生,数据量巨大而形式混杂,在传统的IT技术体系中很难找到良好的方式来有效管理。然而,如果缺失这些很能反映师生用户在校园网的工作、学习和生活方面行为的非结构化数据,那么在数据分析和应用过程中仅能使用有限的业务数据,必然降低了数据利用的价值,局限数据应用和服务手段的发展。

二是业务部门数据利用的意识欠缺

业务部门往往对信息化的意愿不强,主动性不够。对于业务部门来说,利用信息化手段辅助加快业务进程、简化业务过程、保障数据安全,有效提高工作

水平已经达到了自身的预期,而对数据的综合分析,特别是深层次的利用由于过程不直观、成效不明显而动力不足。特别是在当前的行政体系架构下,学校每个部门主要负责自身的业务职能,对有效完成这些业务流程、并进一步提高和完善关乎其自身利益,有较高的积极性;而对于一些数据挖掘、利用产生的效果能影响到其他部门的,由于没有对本部门产生明确的效益而态度消极。

三是数据挖掘、研究层次亟待提升

在很多数据的分析、利用过程中,信息化工作者和业务部门往往满足于对数据的简单分析,即只对能分析出简单的因果关系的数据产生兴趣,例如对一卡通的消费情况进行统计,能很简单地发现学生的住校情况、确认贫困生的消费情况等,这对学工部门的学生工作无疑是起到作用的。然而对于更深层次数据挖掘,例如将学生一卡通消费情况和学习成绩挂钩,却很少能有业务部门能意识到其中存在的相关关系。复旦大学曾经尝试做过研究,发现了使用一卡通消费按时吃早餐的学生人群相对不按时吃早餐的学生人群学习成绩绩点有明显差别。这两者之间虽然没有简单、直观的因果关系,却存在某种必然联系,而该现象没有得到有关部门的充分重视,无法对业务的调整产生良性反馈。

三、大数据的启示

大数据时代的到来为陷入瓶颈的信息化校园数据利用提供了新的出路。大数据技术与应用的一般观点认为:重视数据的全面性而非精确性,将海量数据作为量化事物和行为的一种基本手段,以此可发现事物之间的相关关系,并用于改造流程和优化服务。

众所周知,大数据的特点可以用4个“V”——Volume, Variety, Value, Velocity来描绘,或者说有四个层面:第一,数据体量大,数据量从TB级别跃升到PB级别;第二,数据类型多。如网络运行日志、公开课视频、课程资源图片、无线网或物联网的定位信息等等,除了结构化数据以外,还有规模庞大的非结构化数据。第三,单体价值低,整体价值高。以网络日志为例,连续不间断网络行为中,单一的一个或几个网络行为不能说明任何问题,但大量的、整体的日志则可描绘整个网络访问的趋势。第四,处理效率高。与传统数据处理方式有区别,大数据在处理过程中尽量简化复杂度,不强求数据的连续完整,而尽量高效快捷。

大数据将全面、多源的数据作为一种优化服务的手段纳入视野,对高校信息化服务的发展将产生巨大

的影响。在数字校园建设阶段,学校信息化重视的是业务系统的建设、业务流程的完善和业务数据的积累,这个阶段的数据以结构化为主,采用关系数据库存储,数据利用以数据的共享、展示、统计为主。数据的准确性和安全性是这个阶段数据管理追求的目标。随着高校信息化建设进入智慧校园阶段,用户对服务的要求越来越高,迫使信息化部门对服务的精细化进行深入考虑。智慧校园以个性化服务为目的,势必要求信息化管理者对用户的服务需求进行深层次的发现和挖掘。例如,校园网、物联网、无线网、一卡通、e-Learning 系统、MOOCs、社交平台、生活服务平台、预约系统的日志记录等反映了用户的日常网络使用行为、消费能力、喜好和学习习惯,对于实施“一站式”服务,优化管理环节,以期提升用户使用体验,具有重要意义。相对于传统的数字校园的业务数据,这些能反映“人”的属性的数据数量巨大而且持续不断在产生,大部分是非结构化数据,结构千差万别难以用关系型数据保存和处理。因此以 Hadoop 为代表的许多开源或商业分布式大数据处理架构和 MangoDB、Cassandra 等 NoSQL 数据库进入了校园信息化视野,这类基础应用着重于分布式协同和结构可扩展性的研究,适合于非结构化数据的存储。因此,在智慧校园建设阶段,对数据的保存和利用将以多源、海量的结构化和非结构化数据为主,基础设施的成熟和完善,将为大数据的处理和分析创造基本条件,从此校园信息化建设将从宏观上具备对用户行为和属性的描述能力,这对信息化服务的预测和评估提供了基本依据。

四、从数字化到数据化

大数据并非一个确切的概念,也不是指一种具体的技术,而是指处理的信息量过大、超过一般计算机处理能力的技术和使用思路。在高校信息化领域,大数据的概念刚刚兴起,如何应对大数据的到来是每个高校都会面临的问题。毫无疑问,大数据时代将会涌现针对数据的存储、分析和挖掘的各种技术和应用。然而,对于高校的信息化管理者,在海量数据的获取、汇集、存储问题之外,还应该着重考虑数据化——即采用数据描述事物和现象使之可量化分析的过程。

数字化和数据化大相径庭。在信息化校园中,数字化指的是将职能部门的业务流程转化成业务系统,将业务过程产生的数据转换成电子数据存入数据库的过程。数据化则是指在数字化的基础上,利用数据来对事物和现象进行描述,使之能同时被人和计算机

所理解,量化是数据化的核心。

在信息化校园中,基础数据平台和共享库是存储数据的核心载体,保存的是来自各个管理系统的职能数据,这些数据可为信息化的业务过程直接利用。在数据化过程中,这些只是“原子”数据,通过有效的组织,可以形成服务成效数据化、业务模式数据化、用户行为数据化和用户属性数据化等在信息化服务中管理者比较关心的数据化形式。例如,服务成效数据化可以采用用户使用信息化服务的时间、点击和填写表单的次数、以及服务系统中的评价体系来综合体现,通过服务调整对数据的持续改变来衡量服务优化的成效;业务模式数据化对业务过程进行量化,可依此对业务管理的调整进行预测,例如在财务报销过程中,根据网上预约人数和同期实到人数进行综合评判,业务部门据此预测安排服务人员数量;用户行为数据化可根据用户使用网络系统的日志来分析用户群体在使用系统和网络的过程中对各种资源的停留时间、资源消耗和共同兴趣,教学科研部门可据此来调整教学策略,为用户统一提供相应的协作平台和资源共享服务,而信息化部门则可根据用户行为特征优化网络带宽资源,提高网络利用率和用户体验;用户属性数据化则根据用户自身填报或统计挖掘的数据来对用户的某方面的能力和属性进行描述,以便学校为每个师生提供有针对性的良好服务,例如复旦大学在进行的教师科研教学能力属性的挖掘中,对图书馆历史数据中的论文署名进行了梳理,将教师的论文数据化关联到个人;进一步深入结合教师在 eLearning 网络教学平台的活跃程度、教学评估、论文合作署名等信息,就可以描绘出一个教师的科研能力、科研范围、合作者、教学能力、学生关注度和欢迎程度等,为学校改进教学和科研管理提供依据。

五、大数据创新信息化服务管理

复旦大学在数据化的过程中,逐步完善数据应用机制,建立创新的数据管理和利用体系,包括利用云计算和虚拟化技术建立数据基础设施,探索建立数据量化标准规范,推广个人数据服务以完善数据收集机制,创新数据统计与展示服务业务变革。

在智慧校园建设中充分应用大数据,是提升建设水平、提高服务质量的有效手段。显然,应用大数据并不仅仅是利用云计算、分布式或虚拟化手段完成数据的存储、保管等技术层面的问题,而需要整个校园信息化环境,从业务部门到信息化部门,主动思考,积极探索数据的收集、量化、应用和管理模式。复旦大学在

信息化建设过程中,特别重视数据对信息化对象、业务、过程的刻画,开展了一些有关的特色工作。

1.以“人”为核心建立数据收集和应用机制

师生用户始终是校园信息化服务的核心,围绕“人”来建立数据收集和应用机制,不仅能抓住重点有的放矢,又能简单易行清晰完善。复旦大学建立信息化“一站式”服务平台,整合跨部门业务,统一数据来源,分离后端逻辑和前端界面,使得用户可以专注与友好、统一的服务界面交互,提升用户的应用体验,统一业务数据收集入口;抽取与“人”相关的各类数据,除信息化部门自主维护的系统外,还通过多种方式共享由业务部门独立建设的业务系统中的数据,建立个人数据中心,主要功能包括数据展示、数据填报、数据下载等服务;并规划在个人数据中心基础之上逐步建立、完善校级的统一填报中心,将填报服务与管理流程分离,使所有基础数据对师生可见,减少重复上报,达到数据的高效利用。

2.依托基础数据积累探索数据化标准规范

传统的信息化校园的基础数据标准规范主要适用于学校业务的各类信息编码,统一人员、设备、财务、科研等各项业务的定义和描述,用于指导信息化建设过程中的数据统一。数据化的过程在基础数据规范标准上,对事物与现象进行描述,比如对一个教师而言,应该记录什么信息、何种行为才能展现其科研和教学能力,并无标准,因此开展这方面的探索和研究,具有重要意义。目前,在教师个人的科研和教学活动能力方面,可尝试通过收集 eLearning 网络教学平台中课程资源的教师活跃程度数据、学生参与度数据、教学评估数据、图书馆论文库中的论文作者和影响力匹配数据、学生选课成败历史数据以及其他相关数据来综合评定。

3.完善数据生产与消费循环机制

数据只有在流转中才能体现其价值,有价值的数据才能引起使用者的关注,使数据权威生产者变被动为主动地去维护数据、以保证数据有效性。通过在个人、IT 部门和职能部门之间形成互助互利的数据交流模式,使用户从个人数据中心获取和维护个体属性数据,在填报中心填报业务相关数据;填报需要的基础数据均来自于个人数据中心,最终实现用户无需每次

都要填写重复的个人数据,使得个人数据高度共享,最大限度地减少重复填报次数;业务系统间的数据共享促进数据的流通,减少“信息孤岛”现象的存在。职能部门从数据共享中获得好处(简化工作,提高效率),反过来也促进了管理者在数据利用上的观念转变,推动职能部门使用数据和维护数据的积极性。在数据产生和利用的各个主体之间,建立“个人-院系-学校”自下而上的数据监督与管理机制,形成内需拉动的信息化发展良性循环机制。

4.注重隐私保障数据应用安全

智慧校园的大数据环境下,用户使用校园网的访问行为、一卡通消费、无线定位所处位置等与用户习惯有关的信息进入到业务部门的业务处理范畴,甚至可能通过对数据的挖掘分析出用户的隐私,例如通过一卡通消费可能发现用户的家庭经济状况、通过无线定位可能描绘用户在校园中的行动轨迹等。隐私保护无疑是大数据时代的一个重要课题,是新形态下的数据安全问题。复旦大学在数据利用的隐私保护方面,有如下考虑:首先在数据中心的基础设施上要保障网络安全,部署防火墙和数据审计设备,增强数据安全,杜绝校园网 DMZ 区以外的区域对数据存储系统的直接访问权限;其次,建立数据利用的安全机制,减少数据管理的超级管理员数量,数据开发者和管理者分开,数据由技术部门统一维护和管理,对业务部门的数据分享以隐私消除后的数据为主;第三,在数据应用方面,完善权限管理,以大数据分析预测用户行为为主,以宏观的数据视角支撑业务改革,对于涉及到个人的数据应用做好管理与宣传,减少影响。

参考文献:

- [1]维克托·迈尔-舍恩伯格,肯尼思·库克耶.大数据时代[M].杭州:浙江人民出版社,2013.1.
- [2]赵泽宇,宓詠.云服务引领校园数据中心新发展[J].中国教育网络,2012(10).
- [3]赵泽宇,张凯,宓詠.高校一站式信息化服务[J].科研信息化技术与应用,2012(3).
- [4]大数据时代的大媒体[N].人民日报.2013-01-17.
- [5]大数据概念[N].云财经.2013-05-1.

(编辑:郭桂真)