

用户数据统计挖掘与展示： 以厦门大学图书馆“圃·成长”项目为例

龚晓婷, 肖铮, 陈俊杰, 黄国凡

(厦门大学图书馆, 厦门 361005)

摘要: 大数据时代到来, 人们以新的技术和视角看待数据。图书馆如何应用现有数据, 挖掘和提炼数据价值, 已成为行业关注热点。本文以厦门大学图书馆针对用户数据进行统计挖掘, 通过可视化的方法进行展现为例, 介绍了厦门大学图书馆“圃·成长”项目的设计思路、实施过程和项目展望。

关键词: 数据挖掘 数据可视化 创新服务

一、引言

数据统计和数据挖掘技术已经应用在图书馆用户需求分析^[1]、个性化服务推荐^[2]、读者忠诚度提升^[3]、期刊管理^[4]等方面。近年来随着大数据概念的普及, 人们对于数据的价值以及数据的应用都有了新的认识。大数据给图书馆带来了新的影响和挑战, 同时也给图书馆的服务创新和发展提供了新的机遇^[5]。如何唤醒沉睡在图书馆系统中的用户数据, 提炼它们的价值, 已成为行业关注的热点。

互联网公司对于用户数据的挖掘分析与包装营销给予图书馆新的启发^[6]。上海图书馆2013年春节推出了个人年度阅读账单, 是一份基于读者上一年借阅记录的回顾总结。厦门大学图书馆2013年毕业季, 为毕业生送上了一份独特的毕业礼物: “圃·时光”网站, 收集整理毕业生的借阅历史、入馆记录、论文题目等个人数据, 用清新文艺的图文讲述毕业生的图书馆故事^[7]。

大学生的校园生活, 与图书馆有着密切的关系, 他们在图书馆学习、娱乐、志愿服务, 在书香陪伴下成长。图书馆积累了大量的相关数据, 如学生的借阅记录, 到馆记录, 座位使用记录, 自助文印记录等等。厦门大学图书馆的“圃·成长”项目, 对图书馆各系统中的用户数据进行了清理、关联、统计、挖掘, 进行精心设计和包装, 以信息图表的形式呈现给用户, 使之成为一份读者的数字档案, 记录他们成长的轨迹。

二、“圃·成长”设计思路

“圃·成长”以在校学生为受众对象, 在学年初发布, 目的在于通过数据进行总结和梳理, 让学生看到过去一年自己的到馆次数, 借阅册次、自习时间以及其它服务的使用情况。通过数据让学生更加了解自己, 了解全校用户的整体情况, 更好地引导学生认识到图书馆对个人的影响和作用, 并通过相关数据分析, 推荐图书馆的特色服务, 让学生在将来更好地使用图书馆, 帮助他们更好地成长。

1. 用户数据统计与展示

用户个人数据来自图书馆自动化系统、门禁系统、座位管理系统、自助文印系统。主要数据包括用户上一学年的借阅总量、入馆总数、自习时间、超期滞纳金、自助文印使用量等; 以个人数据为基础, 按照专业、年级、学院的借阅平均数、入馆平均数、自习平均数进行排名。

在图表展示中, 笔者将用户个人数据与同专业、同年级中排名第一的用户数据以及平均数据进行横向对比, 这样用户可以清楚地了解自己与第一名的差距, 以及自己在同专业、同

年级同学中所处的位置。排名靠前，是对用户过去一年努力的肯定。排名靠后，则是善意的提醒，激励他们以年级平均和年级第一作为目标，在新学年能够多来图书馆，多阅读，在下一年的成长账单中“刷”数据。

此外，笔者针对图书馆系统的历史数据进行了统计，生成热门图书列表。其中既有基于上一年度流通记录统计的年度热门图书，也有基于全部历史记录统计出的按专业、按年级的热门图书。系统通过判别读者专业年级信息，将其所借图书与热门图书进行比较，如果热门图书不在其借阅历史记录中，则将这些热门图书推荐给用户；同时结合用户数据分析，推荐图书馆的相关服务。比如对于超期滞纳金较多的用户，推荐图书馆的邮件、短信超期提醒服务；对于预约图书较多的用户，推荐电子图书的服务；对于入馆和自习较多的用户推荐图书馆的书包柜服务等等。

考虑到项目推出后，可能有大一、研一的新生带着好奇心访问，而此类用户并无上一学年统计数据，在登录后，系统将为其推荐新生需要的各种攻略指南，比如如何使用厦门大学的学生邮箱，如何找书，座位系统的使用，二维码在图书馆中的应用等等。

2. 排行榜单的设计

笔者在统计用户个人数据时，分别对各项数据进行了排序处理，由此得出各项数据的用户排行，将最能反映用户与图书馆亲密程度的借阅量、入馆次数、自习时间这三项数据生成榜单，前五名用户分别赋予“书神”、“馆主”、“学霸”称号，既是对获此名号的用户的褒奖，也能激发其它用户设立更高的目标。除了个人排名，还以学院为单位进行了数据统计，计算出各学院同学在借阅量、入馆次数、自习时间三项指标的平均值，由此生成学院 PK 榜——通过此榜单可以看出全校哪个学院的学生最爱阅读，哪个学院的学生最爱来图书馆，哪个学院的学生自习时间最长。

三、“圃·成长”项目的实现

1. 数据整理

从原始数据到项目可用的最终数据须经过数据提取、数据清理、数据统计三阶段。

(1) 数据提取

原始数据散存于图书馆自动化系统、门禁系统、座位管理系统、自助文印系统的 Oracle 和 SQL Server 数据库中。因项目反映的是学生用户上一学年的数据，所以笔者以学号作为用户唯一身份标识，以 2013 年 9 月 1 日到 2014 年 8 月 30 日作为时间区间，从各系统的相关表中抽取可用字段值，保存至临时库中。首先通过自动化系统的用户表，将所有有效的在校生的信息抽取出来，再以此用户数据为基础，抽取其它系统中的相关记录。如从自动化系统中抽取出相关借阅历史记录，从门禁系统中抽取入馆日志记录，从座位系统中抽取选座离座记录，从自助文印系统中抽取历史使用记录等。

(2) 数据清理

数据清理主要以处理无效数据和重复数据为主，例如门禁系统中的记录是以用户单次入馆记录的，为了统计需要，我们需要将用户同一天多次入馆记录合并成一条记录。座位系统中以用户选座和离座的时间点作为日志记录，为了计算用户在座时间，须将日志记录中独立的选座记录和离座记录对应，用离座时间减去选座时间，得出用户单次在座时长。

(3) 数据统计

数据统计的工作主要通过 SQL 的聚集函数和 Excel 的数据透视表功能，分别统计用户的借阅总量、入馆总数、自习时间、常用座位、常去的阅览室、超期滞纳金、自助文印使用量。按照学生类型、专业、年级、学院三个维度对用户进行分组，计算平均值，并得出每个分组中的借阅量、入馆次数、自习时间的排行。

2. 程序设计

程序设计由后端开发与前端实现两部分组成。

(1) 后端开发

因项目主要以数据展示为主，所以后端开发的主要工作是数据库建表，装载各类前期统计数据，页面数据读取。系统交互的程序主要是用户登录和用户留言，在用户登录时生成一个随机字符串作为分享时使用的 URL 参数，也方便了后续的使用统计。

(2) 前端开发

前端开发的任务是数据的可视化展示及响应式页面的实现。结合项目特点，美工设计采用了清新简约的风格，使用四种色彩作为主要设计配色，代表了大学四年，亦代表项目中四项主题数据：我爱阅读、爱馆如家、专属领地、囿的推荐。考虑到用户通过微博微信分享后的受众主要使用移动终端浏览，因此项目采用响应式页面设计，保证移动端的访问体验。笔者选择成熟流行的前端开发框架 Bootstrap，配合页面美工设计，实现基本的页面布局和展示效果。项目的数据展示部分采用开源的前端数据可视化库 ECharts，根据展示内容构建相应的图形样式，如借阅数据通过柱状图展示，自习时间曲线通过面积图展示。图样如图 1 至图 5。



图 1 项目 LOGO



图 2 借阅数据展示图样

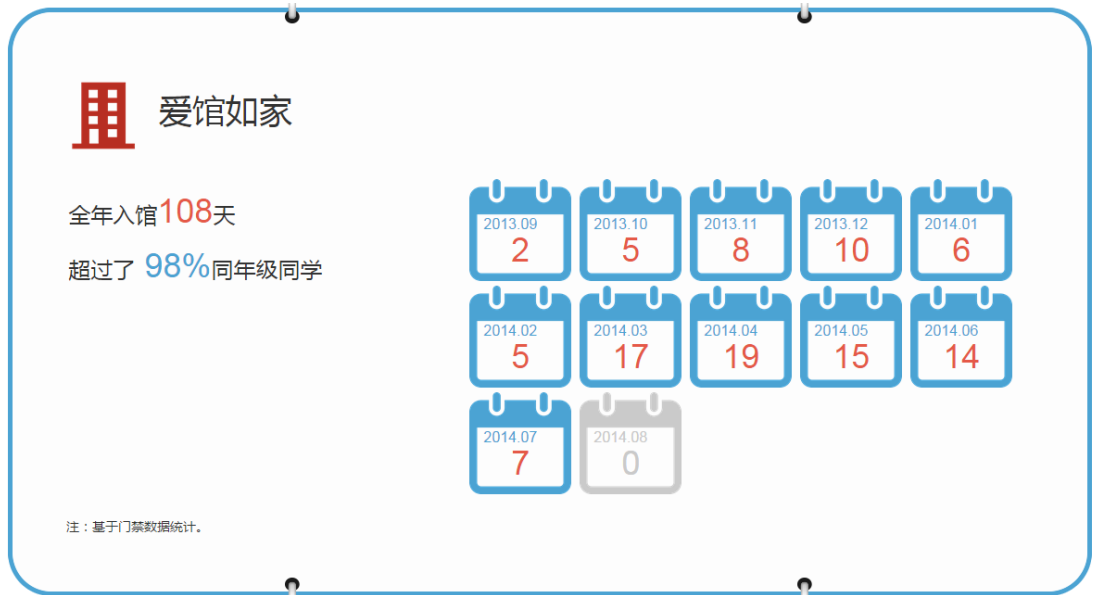


图3 门禁数据展示图样

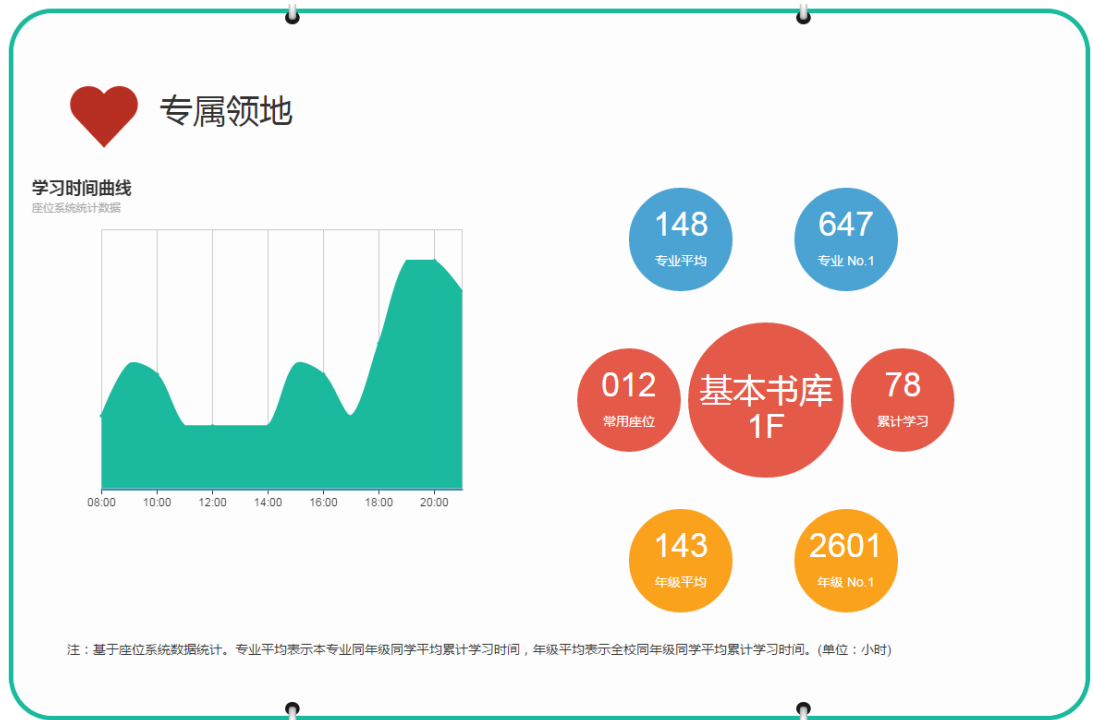


图4 座位系统数据展示图样



图5 个性化推荐图样

3. 宣传推广

项目完成后，通过多渠道跨平台进行宣传推广，充分利用图书馆的官方网站以及微博、微信平台进行营销。笔者在微博开启了“圆·成长”、“院系PK”相关话题，用户可将自己的数据展示页面分享，也可对院系PK榜单进行评论。在不到十天的时间里，笔者通过微信平台推送了六条图文消息，阅读量总计达10912次，而用户在微信朋友圈的分享，亦吸引了大量相关用户登录。此外，为鼓励用户登录与分享，笔者设计了奖励幸运用户的活动，逢整百的登录用户将获得一份珍贵的图书馆纪念礼品。通过以上营销手段，有效地增加了系统使用率，扩大了项目影响力。

4. 效果评估

截止到2015年4月13日，已有4999位同学登录查看过个人的“圆·成长”账单，微博分享229次，相关话题阅读量高达85.2万；通过微博、微信分享链回的总访问量17666次；总浏览量33526次，单用户平均访问次数为3.173次。新华网^[8]和厦门晚报^[9]就此项目专门采访项目组成员及榜单达人，进行了深度报道。

四、结语

“圆·成长”项目尝试使用数据统计、数据挖掘和数据可视化的方法，将读者数据以全新形式呈现给读者，使这份数据账单成为读者成长的印记，成为图书馆送给读者的一份礼物。项目本身也拓展为一次图书馆服务的推广营销活动。未来，笔者将应用新的技术手段，针对用户数据进行更全面深入的挖掘，如分析用户的阅读兴趣、专业特长、学习周期等，并通过用户每年数据的纵向对比，更好地反映用户的成长历程，同时为图书馆的馆藏采选策略、读者服务模式、空间规划设计等提供决策参考。

参考文献：

- [1]李玮平.基于数据挖掘的图书馆读者需求分析[J].图书馆论坛.2004,(03):86-88.
- [2]顾倩.数据挖掘应用于高校图书馆个性化服务的探讨[J].图书馆杂志.2013,(08):63-65.
- [3]姚婉燕.高校图书馆读者忠诚度提升与数据挖掘的融合研究[J].情报科学.2010,(11):1646-1651.
- [4]李继宏.数据挖掘及其在高校图书馆期刊管理中的应用[J].现代情报.2004(07):84-86.
- [5]韩翠峰.大数据时代图书馆的服务创新与发展[J].图书馆.2013(01):121-122.

[6]淘宝 UED.设计提升品牌价值

[EB/OL].[2015-05-21].http://ued.taobao.org/blog/wp-content/uploads/2012/12/1212_淘宝时光机器（设计提升品牌价值）1.pdf.

[7]龚晓婷等.读者数据的挖掘与创意呈现：以“围时光”为例[J].大学图书馆学报,2013(06):92-96.

[8]新华网.厦大晒出年度阅读榜单：看“学霸”是如何炼成的？

[EB/OL].[2015-05-21].http://news.xinhuanet.com/local/2014-12/20/c_1113716434.htm.

[9]厦门晚报.厦大图书馆发布年度排行榜，书神、馆主、学霸纷出炉[EB/OL].

[2015-05-21].http://www.xmnn.cn/dzkb/xmwb/20141215/201412/t20141215_4249713.htm.

联系方式：

（1）通讯地址：福建省厦门市思明南路 422 号厦门大学图书馆

（2）电话：0592-2184973