

校园记忆之图书馆生活 ——基于用户日志挖掘的呈现

汪志莉, 张毅

(华东师范大学图书馆, 上海 200062)

摘要: 通过对高校毕业生利用图书馆数据的挖掘, 分析每个毕业生在校期间到馆及借阅图书情况, 以轻松愉快的方式呈现毕业生个人在校期间的图书馆生活。

关键词: 数据挖掘; 用户日志; 服务创新

中图分类号: G254

DOI: 10.3772/j.issn.1673—2286.2014.10.011

1 引言

近年各大高校都在开展图书馆毕业季主题活动^[1], 包括设计制作“毕业墙”、毕业摄影展、图书捐赠与漂流、毕业生问卷调查、发布毕业生利用图书馆报告等。2013年, 华东师范大学图书馆率先推出基于毕业生利用图书馆数据的“校园记忆之图书馆生活”, 从进馆和借阅两大块入手, 统计到馆情况和借阅情况, 分析读者的借阅倾向, 通过挖掘读者数据助力图书馆服务创新。本文总结两年毕业季活动的算法实现, 希望能对图书馆数据统计分析以及利用数据助力服务创新有所启示。

2 框架设计

毕业季的校园各种情感交集, 各类活动丰富, 图书馆作为大学四年学习生活的主要阵地, 给大家留下点什么呢? 出于让毕业生增加一份对母校的记忆, 增强图书馆的亲和力, 同时宣传图书馆服务的目的, 我们推出“校园记忆之图书馆生活”。设想之初, 我们参考了其他网站, 如新浪微博的“2012, 走你”^[2]、淘宝时光机^[3]、“我的知乎·2012”^[4]及上海图书馆“我的悦读

2012”。在分析本馆数据后, 我们将内容主要分为三块, 即毕业生到馆情况、借阅情况和服务推广。具体框架如图1所示。

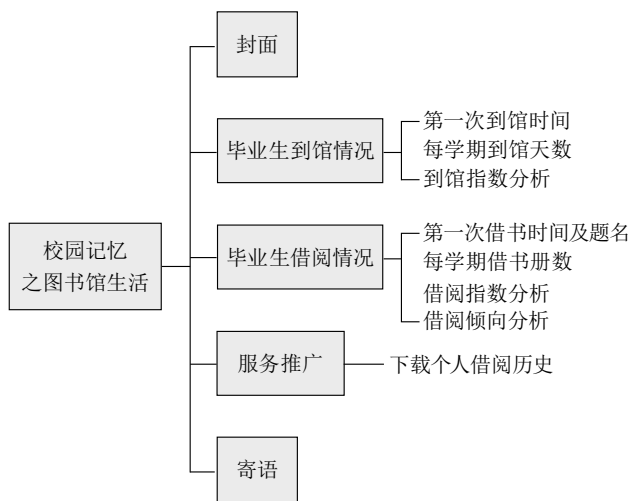


图1 “校园记忆之图书馆生活”框架设计

3 数据来源

按照框架, 逻辑上建立两个库: 毕业生到馆数据库

和借阅历史数据库,如图2所示。

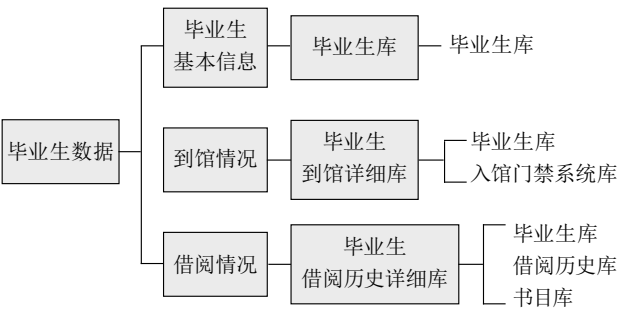


图2 “校园记忆之图书馆生活”数据源

3.1 毕业生库

该数据从图书馆集成系统（ILS）中导出，其中包含学号、唯一标识号、所在学院、读者类型、借书总册数等基本信息。

3.2 毕业生到馆数据库

该数据主要通过“学号”将“毕业生库”和“入馆门禁系统库”关联组成，其中包含毕业生每次到馆的时间和地点（根据通道机编号辨别所在分馆）。

3.3 毕业生借阅历史数据库

该数据通过“读者唯一标识号”将“毕业生库”与

“借阅历史库”（从图书馆系统ILS获得数据）关联，其中借阅历史库包含读者唯一标识号、书目记录号、馆藏记录号、借出时间四个字段，通过“书目记录号”将“借阅历史库”与“书目库”（从图书馆系统ILS获得数据）关联，最终数据包含毕业生基本信息、借阅时间、借阅图书书目记录号、馆藏记录号、题名、索书号等信息。

4 毕业生借阅倾向算法

毕业生借阅倾向是后台数据处理分析的重点和亮点。我们采用与全校或学院借阅倾向的匹配情况来显示毕业生的借阅倾向，全校与学院借阅倾向数据来自年度《馆藏文献利用调查与分析》报告^[5]。通过分析毕业生和样本空间（全校或者院系）的数据，按照中图分类法将读者借阅图书分为22大类，提取借阅最多的前3类与全校或各学院借阅最多的前3类作匹配，将匹配结果分成表1所示的8种情况。

算法中的“完全匹配”指匹配既考虑内容，也考虑出现的次序；“任意匹配”是指匹配时只匹配内容，不考虑次序，此种匹配方法是2013年的改进版本，考虑了样本空间（全校或者学院）借阅最多的前3类的册数差距，本校大部分学院第一类的借阅册数都远大于第二类和第三类，即学院特征比较明显，因此，我们对结果中的前5类作了限制（即对用“标准”和“很像”修饰的毕业生），要求毕业生第一类必须完全匹配样本空间第

表1 毕业生借阅倾向算法思想

序号	毕业生与样本空间借阅前三类比较结果	结果描述
1	如果本人第一类与全校第一类完全匹配，二三类任意匹配	从阅读兴趣看，我是一个标准的XX大学学生
2	如果本人第一类与本学院第一类完全匹配，二三类任意匹配	从阅读兴趣看，我是一个标准的XX学院学生
3	如果本人第一类与其他学院第一类完全匹配，二三类任意匹配	从阅读兴趣看，就读XX学院的我，更像是XX学院的学生
4	如果本人第一类与本学院第一类完全匹配，本人二三类任意一个和本学院第二类匹配	从阅读兴趣看，我很像XX学院学生
5	如果本人第一类与其他学院第一类完全匹配，本人二三类任意一个和其他学院第二类匹配	从阅读兴趣看，就读XX学院的我，很像XX学院的学生
6	如果本人前三类与本学院第一、二类任意匹配（本人第一类和学院第一类不一致）	从阅读兴趣看，我只是有点像XX学院学生
7	如果本人前三类与其他学院第一、二类任意匹配（本人第一类和其他学院第一类不一致）	从阅读兴趣看，就读XX学院的我，有点像XX学院的学生
8	如果匹配不到学院	和我的同学们相比，我的阅读兴趣有点特别

一类(第一、二、三类按照借阅册数递减排序),从发布后毕业生反馈来看,毕业生的借阅倾向描述和实际情况比较吻合。

不同学校可以根据样本空间处理结果来调整算法,如果样本空间前3类的借阅册数相差不大,那么可以将本算法中的完全匹配换成任意匹配,当然本算法还有待进一步完善,力求用语言更准确的描述数据结果。

5 前台数据呈现

5.1 程序实现流程

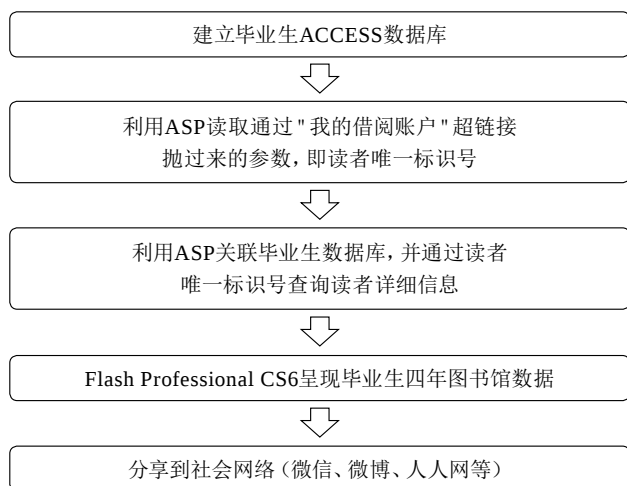


图3 “校园记忆之图书馆生活”程序实现流程

5.2 主要技术应用

FLASH: FLASH^[6]的时间轴上总共包含四个图层,分别是script、mask、imageset、scrollset, scrollset层存放控制flash播放的滚动条, imageset层用来存放背景照片的影片剪辑, mask层存放遮罩图片, script层用来存放as2.0代码。

ASP: ASP负责控制,主要完成三项工作,按照程序运行的顺序依次为:第一,接受从“馆藏目录系统”发送过来的毕业生的唯一标识号;第二,利用毕业生的唯一标识号去“毕业生数据库”中读取毕业生的相关信息;第三,负责与FLASH通信。

分享: 分享功能采用百度分享(<http://share.baidu.com/code>)来实现,百度分享功能可以将网页分享到新浪微博、腾讯微博、QQ空间、人人网、豆瓣等等,还可

以对网站的浏览次数、分享次数作统计。

6 结果展示

考虑个人隐私,查看及下载需要从“我的借阅账户”登录,截图如图4所示。

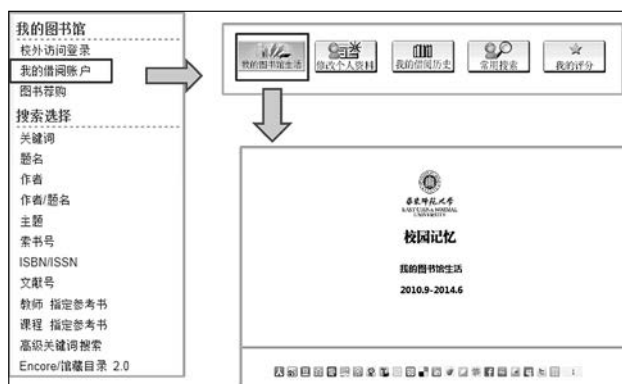


图4 “校园记忆之图书馆生活”登陆入口

到馆部分截图如图5所示。

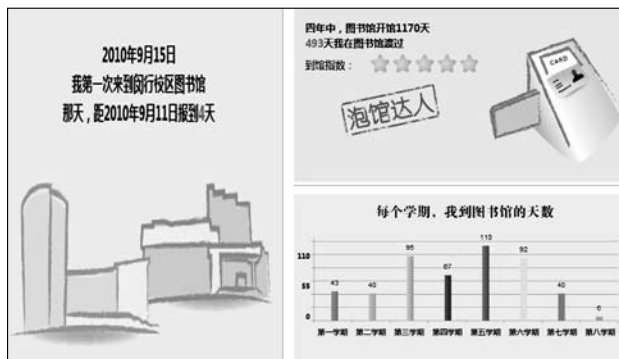


图5 “校园记忆之图书馆生活”到馆情况展示

借阅部分截图如图6所示。

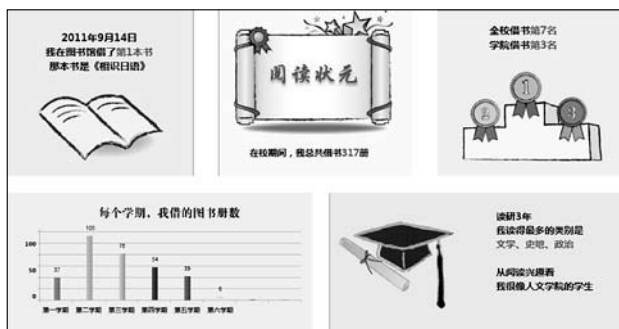


图6 “校园记忆之图书馆生活”借阅情况展示

7 结语

“校园记忆之图书馆生活”从个人“记忆”角度以讲故事的方式娓娓道来,全过程采用第一人称,而非常用的第二人称,语言表达诙谐,有点“好玩”、“有意思”的元素,文艺范中带着温馨熟悉的感觉,靠数据说话,但不堆砌数据,对本人的数据精确化,起到保存“记忆”的作用,但与他人比较的数字尽量模糊化,只以不完全明晰的等级、归类表示,例如到馆等级模拟QQ等级,用满五颗星的方式来呈现,借阅等级则按照中国古代科举考试名次来划分,既含有现代元素,又有浓厚的文化底蕴,初次看到令读者有耳目一新的感觉。借阅倾向归类,更是求同存异,读者反映良好。

华东师范大学图书馆目前采用FLASH编程读取数据库方式,而数据的表现形式有多种,如网页方式、动画方式、相册方式等等,对于服务的推出可以采用嵌

入读者个人账号页面或者主动推送读者电子邮件等方式。我们将继续挖掘图书馆各类数据,积极学习各类流行的共享方式,使数据更多地被读者利用和参考。

参考文献

- [1] 于静,孙媛媛,赵敏.高校图书馆毕业季主题活动的调研、实践与思考[J].大学图书馆学报,2013(6):88-91.
- [2] 走起2012 [EB/OL]. [2014-09-27]. <http://place.weibo.com/z/love2012>.
- [3] 淘宝时光机[EB/OL]. [2014-09-27]. <http://me.taobao.com>.
- [4] 我的知乎•2012 [EB/OL]. [2014-09-27]. <http://www.zhihu.com/topic/19801232>.
- [5] 2013年度华东师范大学图书馆馆藏文献利用调查与分析[EB/OL]. [2014-09-27]. http://www.lib.ecnu.edu.cn/about/lib_annual/.
- [6] 阎瑞华,张静.Flash CS5中文版基础教程[M].北京:人民邮电出版社,2014:132-146.

作者简介

汪志莉,女,1983年生,硕士,华东师范大学图书馆馆员,E-mail: zlwang@library.ecnu.edu.cn。
张毅,男,1986年生,硕士,华东师范大学图书馆馆员。

Presenting Library Journey of Campus Memory: Based on User Log Mining

WANG ZhiLi, ZHANG Yi
(Library of East China Normal University, Shanghai 200062, China)

Abstract: By data mining for the graduates using academic library, the information of accessing library and circulation for each graduate in the school is analyzed, and the library life of individual graduates is presented by lively style.

Keywords: Data mining; User log; Service innovation

(收稿日期: 2014-10-09)