



华东师范大学图书馆

用数据说话：

数据分析驱动下的馆藏利用率提升实践

系统部 汪志莉



大纲

本馆现状

- 借阅情况
- 馆藏增长情况
- 借阅倾向分析

大数据环境下的方法研究

- 个性化图书精准推荐
- 特色馆藏揭示

传统方法

- 阅读推广
- 馆藏布局调整
- 修改借阅规则



引言

数据工作已有成果简介

——在此基础上更深化服务与深入了解用户行为

白皮书



借阅排行榜



毕业季

为研究生院、
教务处推荐数
据



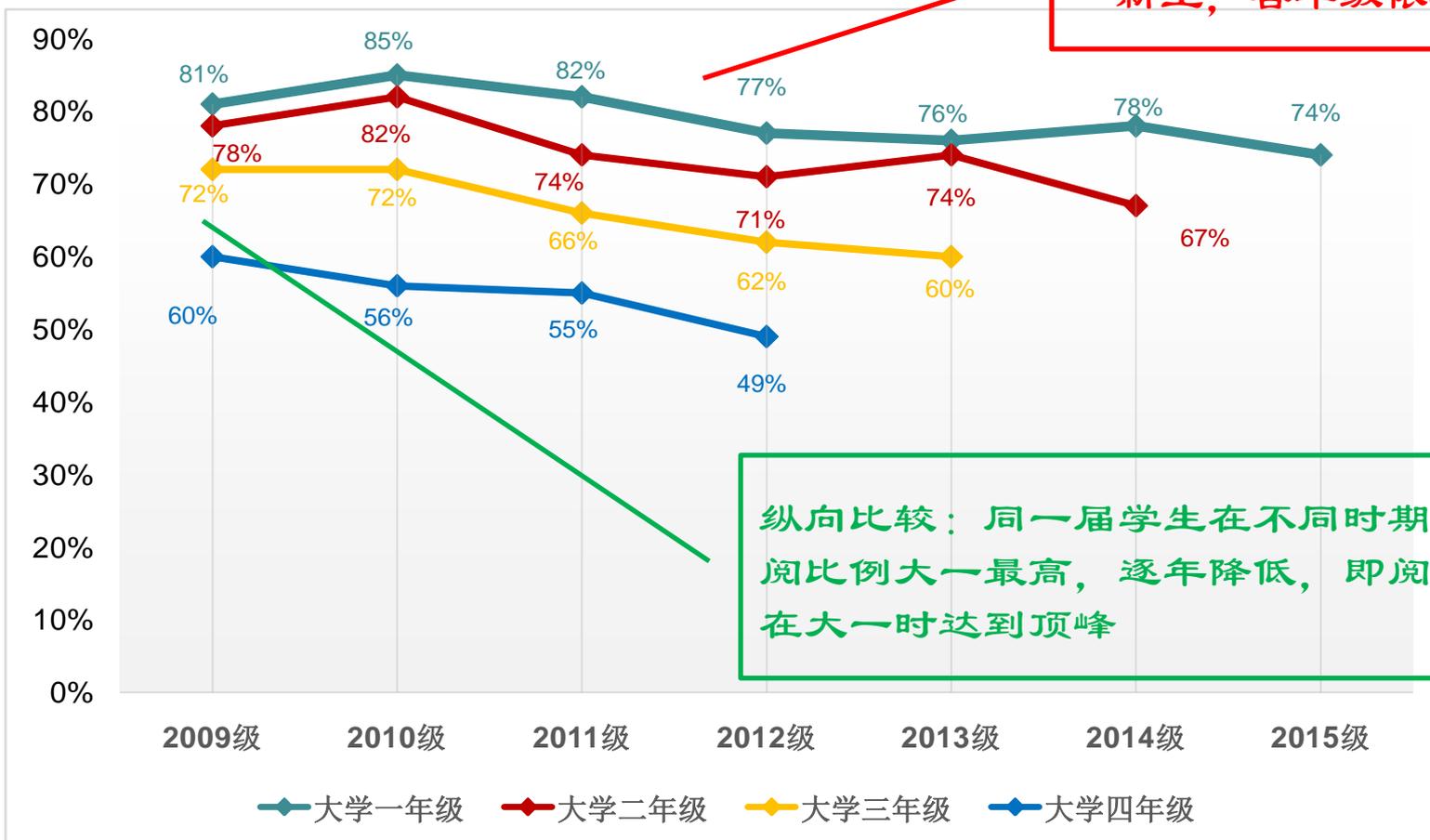
公管院数据请求等等...

统计工作的必要性以及扎实的数据基础可能产生的衍生服务



本馆现状 —— 借阅情况

不同年级学生大学四年借书百分比



横向比较：同一年不同年级的学生，借阅百分比最高的是大一新生，各年级依次降低

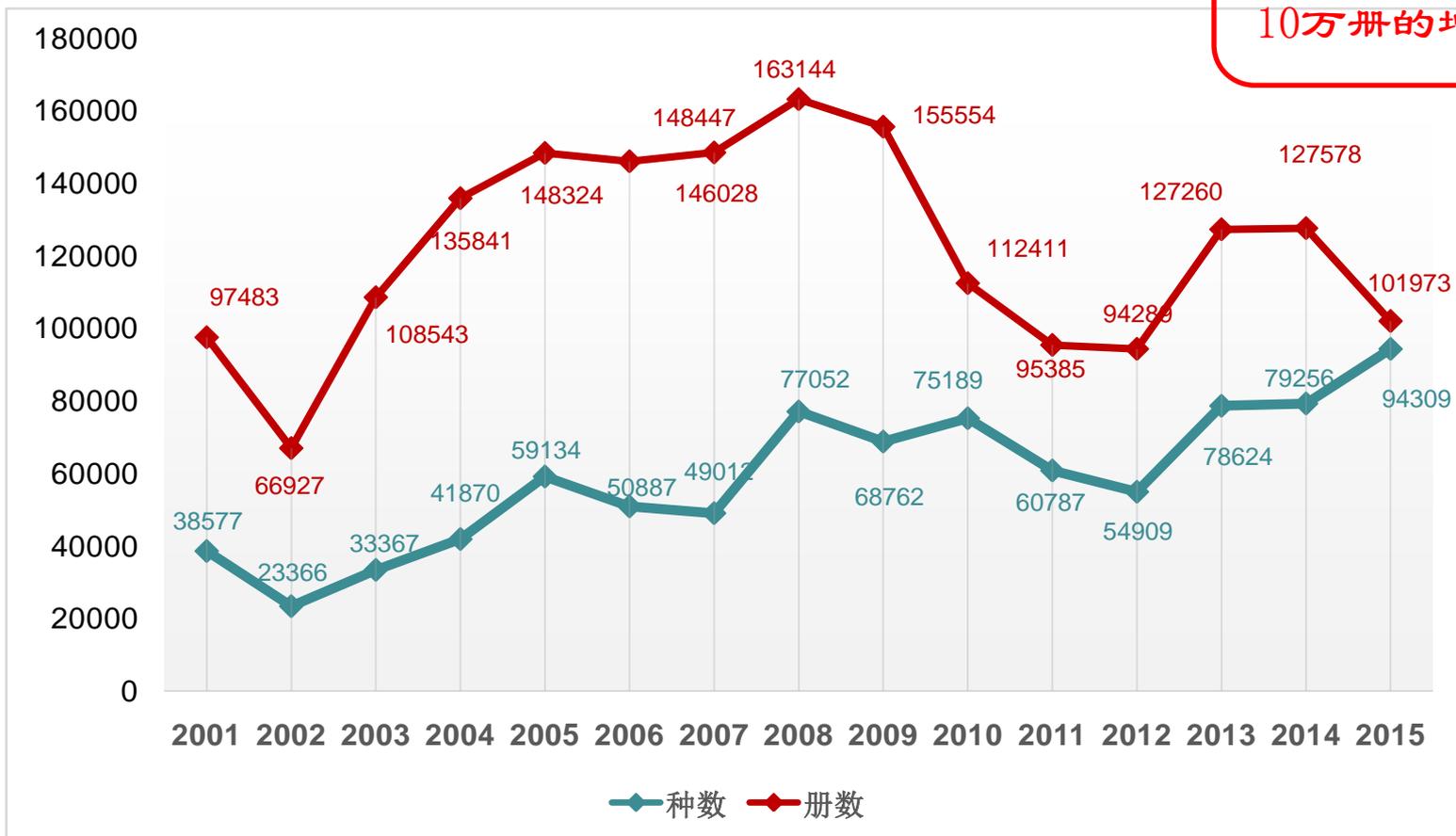
纵向比较：同一届学生在不同时期，其借阅比例大一最高，逐年降低，即阅读兴趣在大一时达到顶峰



本馆现状 —— 馆藏增长情况

2001-2015各年新增馆藏种数及册数比较

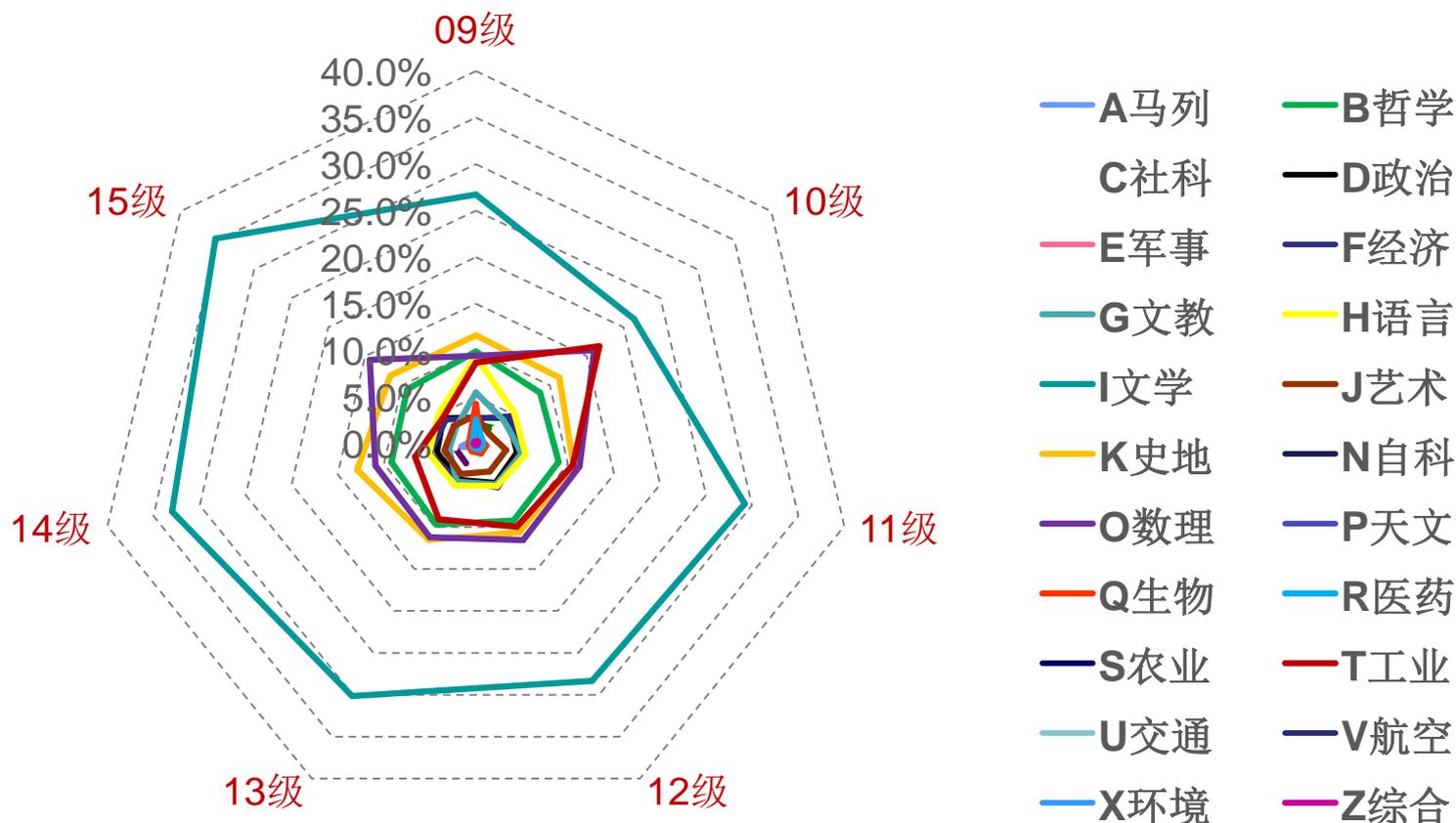
从2003年起，馆藏每年基本保持10万册的增长量





本馆现状 —— 借阅倾向分析

— 以各级本科生22大类借阅分布为例



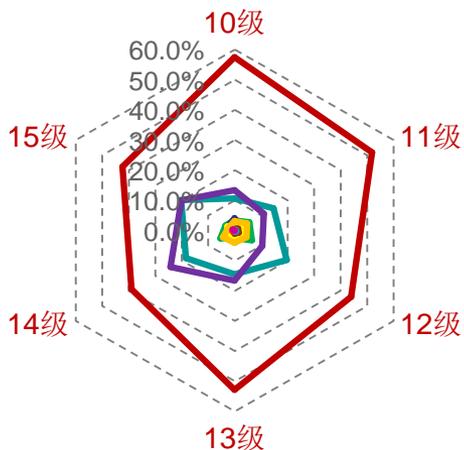


本馆现状 —— 借阅倾向分析

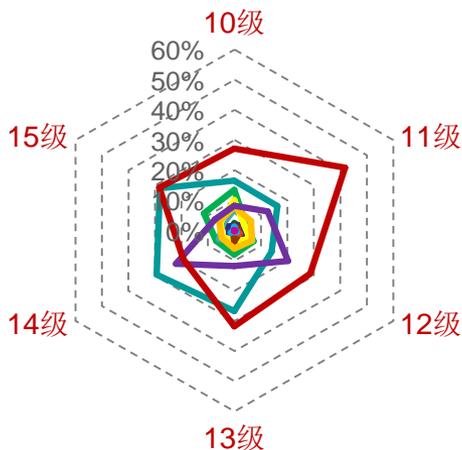
— 举例三大学院各年级借阅分布

- A马列
- B哲学
- C社科
- D政治
- E军事
- F经济
- G文教
- H语言
- I文学
- J艺术
- K史地
- N自科
- O数理
- P天文
- Q生物
- R医药
- S农业
- T工业
- U交通
- V航空
- X环境
- Z综合

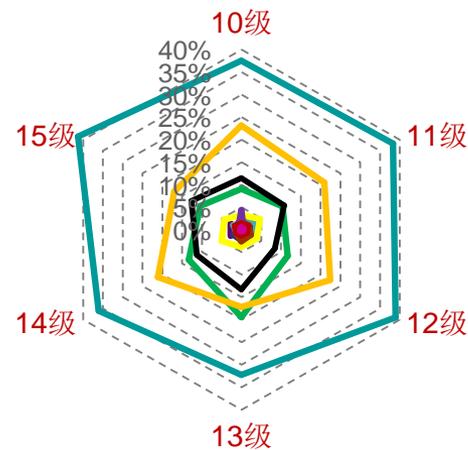
计算机科学与软件工程学院



教育科学学院



人文学院



图像反映借阅倾向各届学生各不相同，由此需要我们借助大数据技术、数据挖掘进行更加精确、针对性、智能化的主动推送。



缘起：

✓ 借阅量逐年下滑的触动，考虑寻求提升读者借阅量的

方法

✓ 互联网思维影响，以用户为中心的思考，丰富**我的图**

书馆内容

✓ 让推荐更有针对性、主动性、智能化



算法确认：

✓ 协同过滤

✓ Slop-one

✓ SVD (奇异值分解)

✓ Knn

✓ 聚类

.....



原理：

- ✓ 将一个用户所借阅偏好作为一个向量来计算用户之间的相似度
- ✓ 确定群体数 ($K=20$)
- ✓ 根据群体读者**相似度权重**以及所借书的偏好，预测当前用户可能的需求，计算推荐内容



大数据环境下的方法研究 — 个性化图书精准推荐

流程：





大数据环境下的方法研究 — 个性化图书精准推荐

推荐书目嵌入个人借阅账户 (My Library) :

馆藏目录

华东师范大学图书馆

[首页](#) | [Eng](#) | [图书馆主页](#)

刘丹
闵行区东川路 500 号
上海
上海
200240
dliu@library.ecnu.edu.cn

我的图书馆

退出

返回我的记录

重新搜索

图书荐购

搜索选择

关键词

题名

作者

作者题名

主题

索书号

ISBN/ISSN

分类号

文献号

教师指定参考书

课程指定参考书

高级关键词搜索

Encore 馆藏目录 2.0

手机 OPAC

使用指南

帮助中心

外借排行

新书通告

文献借阅规定

图书分类规则

与你兴趣相同读者也喜欢:

2 借出论框					
续借	题名	评分	条形码	状态	索书号
<input type="checkbox"/>	LINUX技术手册 / Ellen Siever ... [美] 著; 姜云川译	☆☆☆☆☆	10009834958943	到期 15-05-02 已读 2 次	TP316.85 X119 2013
<input type="checkbox"/>	Linux Shell脚本攻略 = Linux Shell Scripting Cookbook / (印) Sarath Lakshman 著; 门倍译 eng	☆☆☆☆☆	10009832821762	到期 15-05-16 已读 1 次	TP316.81 L1118 2011

© 2010 华东师范大学图书馆 版权所有



实施效果：

通过分析系统日志，统计时间段为2015.3.9-2016.4.22的推荐使用情况，在所有的登录读者中，约17%的读者查看了推荐，约4.4%的读者借阅了推荐的图书。对推荐图书产生了查看行为的读者，平均查看推荐数为3-4本，产生了借阅行为的读者借阅量为1-2本。但也有读者借阅了7本图书。

统计量	数值
统计时间	2015.3.9-2016.4.22
登录读者 (人)	23741
查看推荐读者 (人)	4030
查看推荐读者比率	17.0%
借阅推荐读者 (人)	181
借阅推荐成功比率	4.4%
查看推荐最高值 (本)	120
借阅推荐最高值 (本)	7



大数据环境下的方法研究 —— 特色馆藏揭示

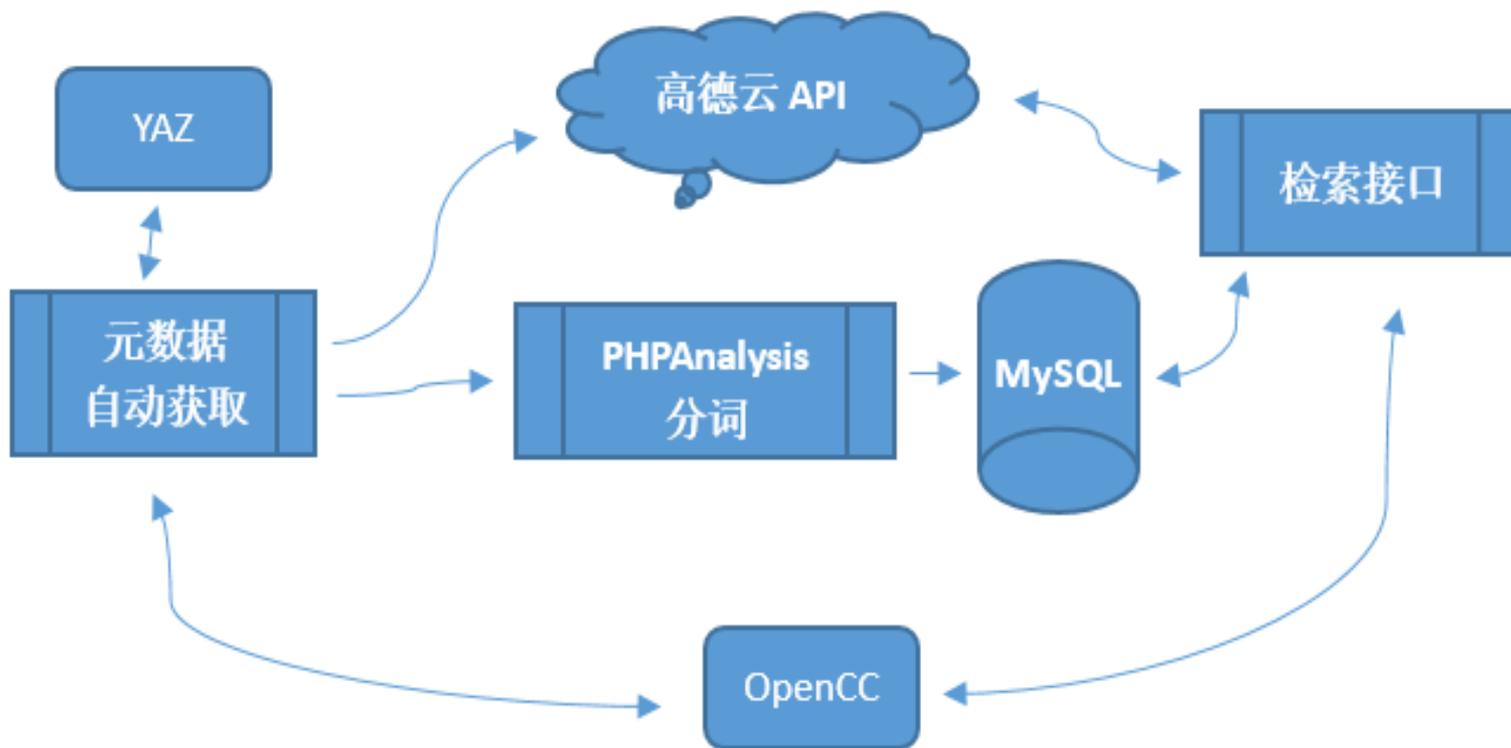
缘起：

- ✓ 图书馆在馆藏特色资源数据库建设方面不断加大投入，具有特殊价值的独有数据库大量形成
- ✓ 弥补目前馆藏OPAC、部分专用方志资源平台在方志的检索、资源发现系统在技术上缺乏针对性、新颖性的缺点
- ✓ 从发现系统的角度，构建专有的面向方志的发现平台，通过引入新的GIS、标签云等技术，提供更多与资源特性相关的发现手段，提高资源的可发现性



大数据环境下的方法研究 —— 特色馆藏揭示

流程：





大数据环境下的方法研究 —— 特色馆藏揭示

实施效果：



可视化地图搜索





大数据环境下的方法研究 —— 特色馆藏揭示



实施效果：

标签云

按关键词缩小查询范围：

工业 松江 嘉定县 卷首 馆藏 建校 校史 真如 中国 商业 上海市 月浦 青浦 通志 第一 松江县 金山 卷末 图书 南汇 地名 续编 辞书 1卷 地方志 方志 文化 交通大学 江苏 水利 出版社 县志 第二 二一 嘉定教育 一七 钱门塘 上海 大学 金山县

按地域缩小查询范围：

闵行 嘉定县 徐汇 江苏 崇明县 江西 宝山县 川沙县 普陀 金山县 上海 云间 奉贤县 松江府 青浦县 南汇县 松江 华亭县 松江县 青浦 奉贤 河南 南汇 嘉定 云南 浦东新区 浙江 松江区 宝山 金山 上海县

按年代缩小查询范围：

明代 雍正 天启 崇祯 乾隆 嘉靖 咸丰 正德 同治 宣统 万历 绍熙 民国 清代 道光 弘治 光绪 宋代 顺治 嘉庆 康熙

按关键词缩小查询范围：

工业 松江 嘉定县 卷首 馆藏 建校 校史 真

按地域缩小查询范围：

闵行 嘉定县 徐汇 江苏 崇明县 江西 宝山县

按年代缩小查询范围：

明代 雍正 天启 崇祯 乾隆 嘉靖 咸丰 正德 同治 宣统 万历 绍熙 民国 清代 道光 弘治 光绪 宋代 顺治 嘉庆 康熙



传统方法 —— 阅读推广

缘起：

- ✓ 图书馆阅读推广活动风生云起，借助4.23世界读书日，每年4月份的“阅读月”活动大多办成了有影响的常规品牌活动
- ✓ 加大对纸本资源的推广力度
- ✓ 分主题、分层次推荐图书馆的各类图书，开拓读者的视野，提高馆藏的利用率



传统方法 —— 阅读推广

方法实施：

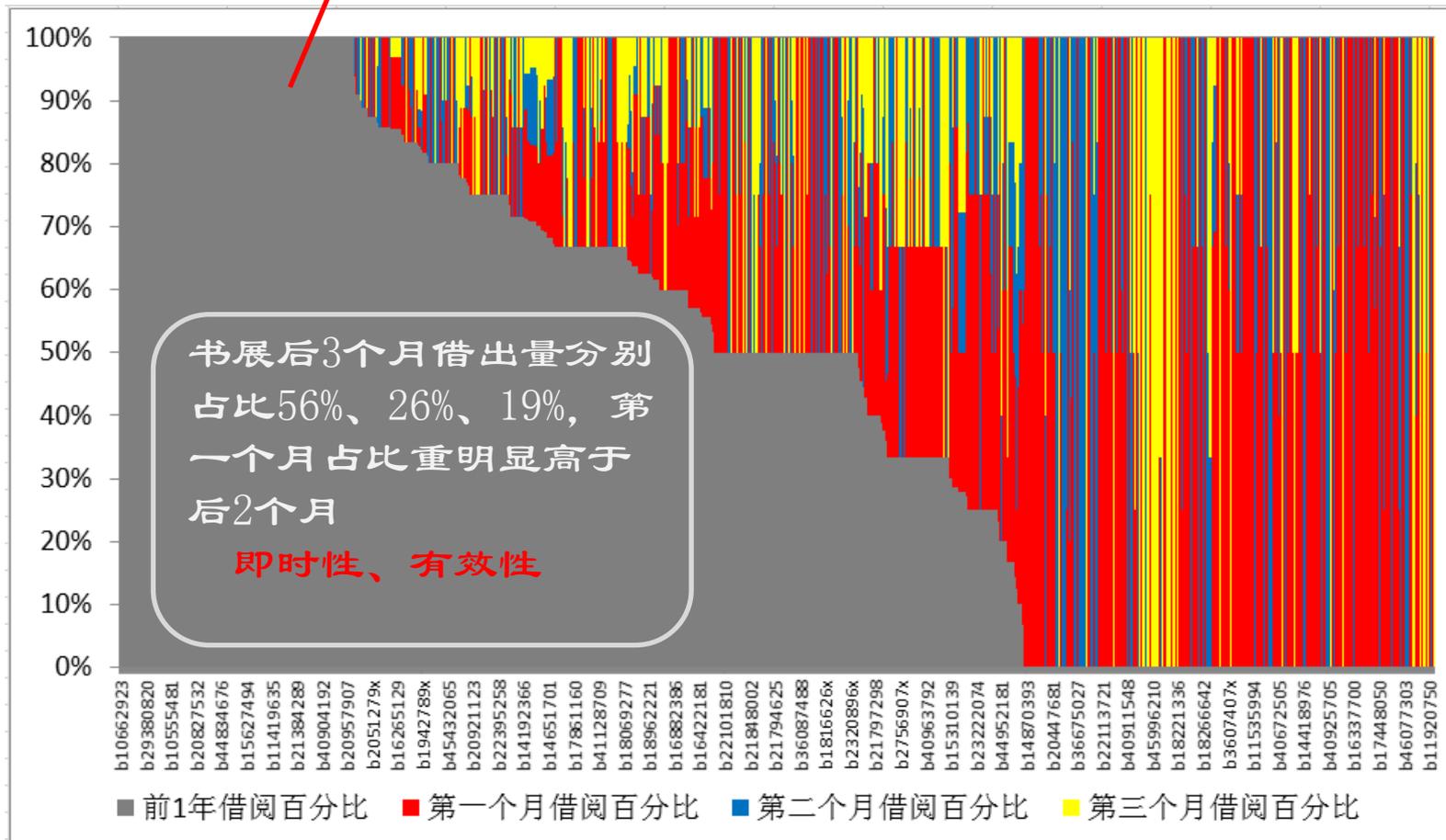
2014年开始陆续推出书展活动，2014年总共推出13期，2015年总共推出19期，**书展的效果到底如何？**并没有来验证和核实，抓取书展时间在**2015.1.1-2015.12.31**的所有馆藏记录号码，根据每本馆藏的书展时间，分时间段统计借阅册数进行比较分析，主要分为4个时间段，分别是书展前1年、书展推出第1个月、书展推出第2个月、书展推出第3个月，用统计数据对比分析书展效果。



传统方法 —— 阅读推广

书展后3个月的借阅总量 \geq 前1年的借阅总量的有859种图书（有效比较数据为1566种），占比55%。

实施效果：





传统方法 —— 馆藏布局调整

方法实施：

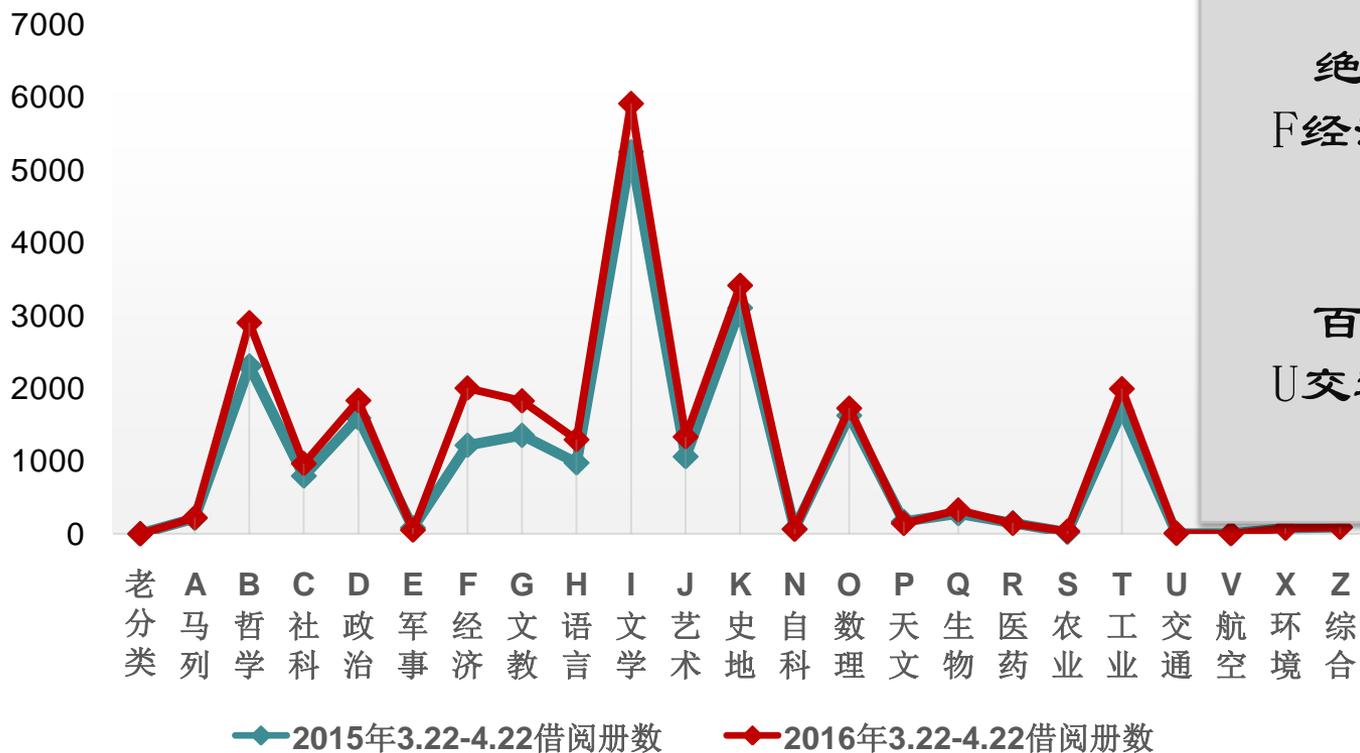
为了进一步实施大开间统一借阅的馆藏布局思想，以及考虑尽量缩短读者在借阅中找书的时间，2016年2-3月，华东师范大学图书馆进行了馆藏布局调整，将闵行校区图书馆四楼、五楼图书进行合并后重新上架，打破原来新书时间人为界定，改变A-Z大类图书根据进馆时间分别放置两层的现状。合并后的五楼布置A-H大类图书，四楼布置I-Z大类图书，并将借阅期限统一为一个月（博士、教师为两个月）。



传统方法 —— 馆藏布局调整

实施效果：

馆藏布局调整区2015-2016年同时段借阅册数比较



排行榜：

绝对值增加最多：
F经济、B哲学、I文学

百分比增加最多：
U交通、S农业、F经济



传统方法 —— 修改借阅规则

方法实施：

为了提高图书的使用率，近10年内，本馆进行了3次大规模的借阅规则调整，分别是2009年、2012年、2015年，每次都对在校科研人员的借阅、续借、预约等权限进行了大幅度的调整。单次借阅册数调整从2009年以前的10本扩大到50本；2009年图书可续借最大次数调整为5次，2015年在2012年的基础上，增加可预约5本的新规则。具体实施规则如下表。

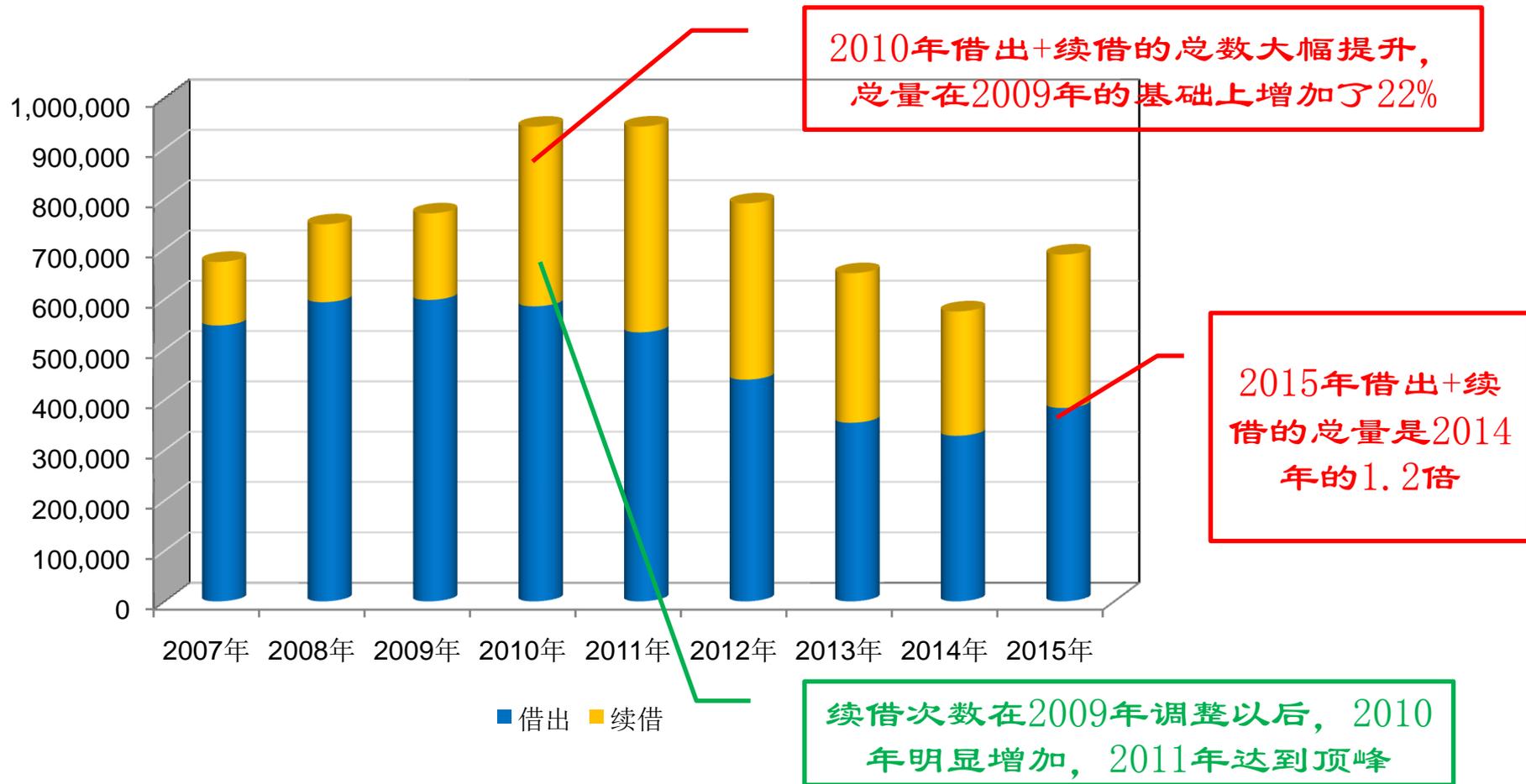
各类读者各次最大借阅册数调整

	2009年	2012年	2015年
教职工、特聘教师、访问学者、博士、硕士	10->20	20->25	25->50
本科生四年级	5->10	10->20	20->30
本科一~三年级	5->10	10->15	15->20



传统方法 —— 修改借阅规则

实施效果： 2007-2015年馆藏外借及续借情况统计





总结

从客观数据的结果来看，修改馆藏布局调整、借阅规则、阅读推广等方法见效较快，如续借规则改变实施后的1年，其续借量是原来的一倍，又如书展数据，3个月的借阅量超过1年的借阅量的占55%，这都体现了使用这些方法提高馆藏借阅量的即时性和有效性；相比较而言，个性化图书精准推荐成功率较前3种来说，相对较低，其一，数据本身受到邱均平等研究印证的数据稀疏性条件下推荐效率的客观情况的影响，其二，和读者个体的喜好有关，需要时间和宣传来逐步推广，但从服务的角度，推荐系统为读者提供一种新的资源发现途径，而读者对推荐也产生了兴趣，并产生借阅行为以及读者的互动也表明推荐系统受到了读者的肯定；新的特色资源发现系统，将标签云、GIS等技术融入系统的设计当中，为用户提供了更加友好的界面，从数据底层到界面都进行了彻底的更新，由于该应用上线的时间较短，目前还没有数据进行证实，但模仿目前主流发现系统的一站式搜索，已经成为目前以及今后开发新系统的必然趋势。提高馆藏资源任重而道远，需要我们不断摸索、实践、总结、改善，需要我们做一个有心人，观察、记录读者阅读习惯、行为、喜好，为各类活动策划提出更好的意见和建议，需要我们不断学习更多数据挖掘、大数据等技术，在技术上跟上时代的潮流，利用各种新颖的工具助力图书馆的各项服务更加与时俱进。



华东师范大学图书馆

谢谢！