



## CASHL 前瞻性课题

### 基于客观数据的 CASHL 用户科研信息和服务需求实证研究

——以南京大学 CASHL 用户为例

CASHL 课题研究小组  
南京大学图书馆  
2016 年



## 本项目主要参加人员

史梅 南京大学图书馆副馆长  
袁晓园 南京大学图书馆 本项目负责人  
慎月梅 南京大学图书馆  
刘熙春 南京大学图书馆

---

---

## 说明

---

---

1. 本报告为南京大学用户在使用开世览文资源与服务中进行调查与分析的课题成果，主要内容基于邮件随访、问卷调查、深度访谈等实证研究方法的基础上，对南京大学随机抽样用户的客观数据和随访数据进行比较和分析。

2. 本报告结构上分为“研究背景”、“研究技术路线”、“研究方法”、“研究样本选择依据和用户的一般特征”、“用户申请的定量分析”、“有科研产出和无科研产出用户的特征和申请特征分析”、“CASHL 申请资源和用户研究方向之间的文本相关度”、“用户申请的时间变化特征”、“用户申请量、花费和科研产出的关系”、“被拒绝的申请分析”、“结论”、“致谢”、“附录”、“参考文献”等部分。其中，报告正文 22 页，附录共 2 页。

3. 本报告所有信息均通过合法手段，利用公开信息源采集而来，未侵犯国家秘密以及商业秘密。

4. 本报告在完成过程中，参考了国内外有关文献，所有分析均为本研究小组独立完成，未侵犯他人知识产权。

## 报告导读

---

本报告拟从用户体验入手，对用户使用 CASHL 的情况进行量化处理，将用户使用或科研产出与 CASHL 服务之间的关连做定量分析，从而获得改进 CASHL 学术资源服务重要的影响因素。本项课题研究基于南京大学的开世览文用户，采用文献调研法、调查法、数据获取程序和统计分析方法，收集与统计南京大学 135 名开世览文样本用户的一般特征和申请特征，分析有科研产出用户、无科研产出用户的一般特征和申请特征；分析有科研产出用户的研究方向与申请资源的文本相关性；对用户的申请量、花费和科研产出的关系做深入分析；进一步分析用户申请的时间变化特征和被拒绝的申请特征，通过上述的分析和比较，尝试获得提升 CASHL 资源服务的重要影响因素。

本报告的正文部分从“研究背景”、“研究技术路线”、“研究方法”、“研究样本选择依据和用户的一般特征”、“用户申请的定量分析”、“有科研产出和无科研产出用户的特征和申请特征分析”、“CASHL 申请资源和用户研究方向之间的文本相关度”、“用户申请的时间变化特征”、“用户申请量、花费和科研产出的关系”、“被拒绝的申请分析”、“讨论”“结论”“致谢”14 个部分综合介绍整个课题的研究内容：

# 目录

说明 .....	二
报告导读 .....	三
<b>1 研究背景.....</b>	<b>5</b>
<b>2 研究技术路线.....</b>	<b>6</b>
<b>3 研究方法.....</b>	<b>6</b>
3.1 文献调研法.....	6
3.2 调查法.....	7
3.3 数据获取程序.....	7
3.4 统计分析法.....	7
<b>4 研究样本选择依据.....</b>	<b>7</b>
<b>5 用户的一般特征.....</b>	<b>7</b>
5.1 性别分析.....	7
5.2 所在学科分析.....	7
5.3 学位和职称分布分析.....	8
<b>6 用户申请的定量分析.....</b>	<b>9</b>
6.1 性别申请量分析.....	9
6.2 申请的学科特征分析.....	9
6.3 申请的学位和职称分析.....	10
6.4 用户申请量和申请次数分析.....	11
<b>7 有科研产出和无科研产出用户特征和申请特征分析.....</b>	<b>11</b>
7.1 有科研产出用户的一般特征.....	12
7.2 无科研产出用户的一般特征.....	14
7.3 有产出和无产出的申请量和用户对比.....	15
<b>8 CASHL 申请资源和用户研究方向之间的文本相关度.....</b>	<b>16</b>
<b>9 用户申请的时间变化特征.....</b>	<b>17</b>
<b>10 用户申请量、花费和科研产出的关系.....</b>	<b>18</b>
<b>11 被拒绝的申请分析.....</b>	<b>19</b>
<b>12 讨论.....</b>	<b>20</b>
<b>13 结论.....</b>	<b>21</b>
<b>14 致谢.....</b>	<b>22</b>
附录 .....	22
表格目录.....	23
图表目录.....	24

## 1 研究背景

近 10 年，在中国高校科研评价体系与国际学术合作环境的联合驱动下，我国人文社会科学学者不断努力，C9 高校中已经有 7 所高校的社会科学进入了世界 ESI<sup>[1]</sup>前 1% 的学科排名，南京大学在其列。为了推动中国人文社科高速发展，不断满足科研人员对外文数字资源日益增长的需求，改善外文数字资源整合平台的匮乏，改进资源获取途径，2004 年 3 月，在国家教育部的领导下，中国高校人文社会科学文献中心（China Academic Social Sciences and Humanities Library，简称 CASHL，又名开世览文）<sup>[2]</sup>正式启动，开世览文是高校哲学社会科学教学和研究建设的文献保障服务体系，而且是全国性的唯一的人文社会科学文献收藏和服务中心。

从 2004 年至今，作为华东北区域中心，南京大学图书馆在提升 CASHL 学术影响力，提高文献资源满足率和尝试探索本校用户的科研信息需求方面，做出了诸多的尝试和努力。2015 年，新重组的馆际互借成员多次走进全校各大人文社科学院，张贴海报，采用新媒体平台发布微信和微博，通过宣讲的形式吸引新用户的加入和提交文献申请；2015 年 6 月南京大学馆际互借小组走入江宁大学城，为江宁 10 多所兄弟院校图书馆同仁介绍和推荐 CASHL；小组采用邮件，面对面访谈等多种形式与 CASHL 用户深度交流，获得重要信息。本研究将本校用户作为研究对象，尝试揭示用户明确的科研信息需求，从而提供更好的服务。

## 2 研究技术路线

本研究拟从用户体验入手，对用户使用 CASHL 的情况进行量化处理，将用户使用或科研产出与 CASHL 服务之间的关连做定量分析，从而获得改进 CASHL 学术资源服务重要的影响因素。

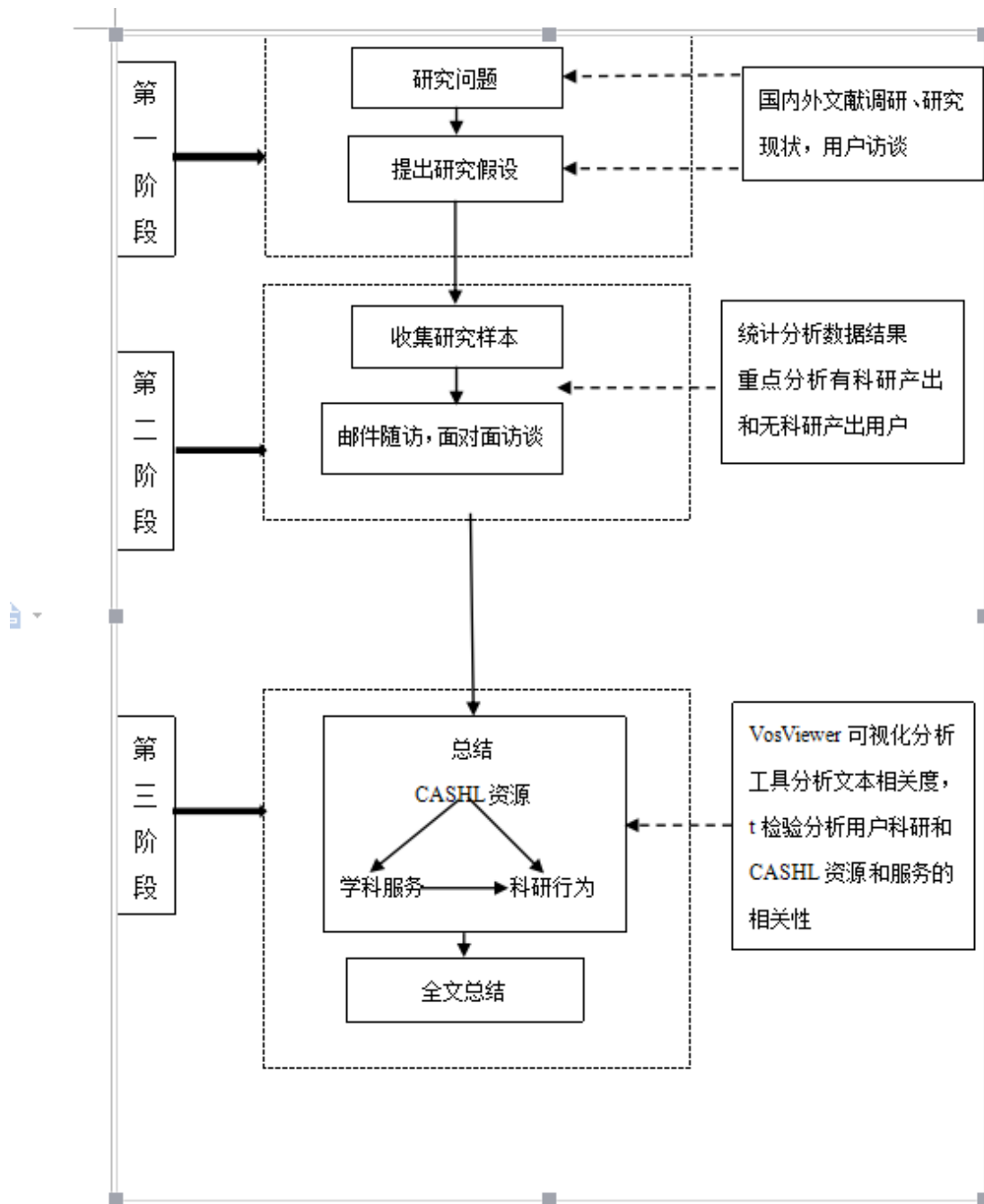


图 1 研究设计路线图

### 3 研究方法

#### 3.1 文献调研法<sup>[3]</sup>。

本研究收集有关外文数字资源利用、用户需求满意度、持续使用影响因素等相关的研究文献，并且跟踪相关领域的发展趋势和最新研究动态，试图把握用户行为、需求与资源、服务提供三者之间相关性的发展脉络及进展，发现研究中存在的问题，尝试提出研究问题和研究假设。

#### 3.2 调查法<sup>[4]</sup>。

本研究为了获取南京大学 CASHL 用户的相关资料，采用邮件随访、问卷调查和深度访谈。

### 3.3 数据获取程序

采用 php 程序，模拟浏览器抓取 CASHL 工作系统的网页数据，并存放在可检索的页面上，<http://114.212.7.152/cashl/index.php>(须在南京大学校园网内登录)可供分析。

### 3.4 统计分析法。

采用 VosViewer<sup>[5]</sup>可视化分析工具分析 CASHL 资源和用户研究方向之间的文本相关度。采用 SPSS 对 CASHL 申请量和申请类型对用户的科研产出是否存在影响作配对 t 检验，同时采用 Spearman 作相关性分析。

## 4 研究样本选择依据

2015 年 3 月，南京大学图书馆对馆际互借进行了重组，小组由新成员组成。为了获取有效的研究对象，在 CASHL 工作系统中随机选择 2015 年 01 月 01 日-2015 年 12 月 31 日时间段内 CASHL 有效账户用户，共收集了 135 人，这些用户均属于在校学生或在职教师，是有效研究对象，可作进一步的问卷调查和深度访谈。

## 5 用户的一般特征

### 5.1 性别分析

男性 58 人，女性 77 人。

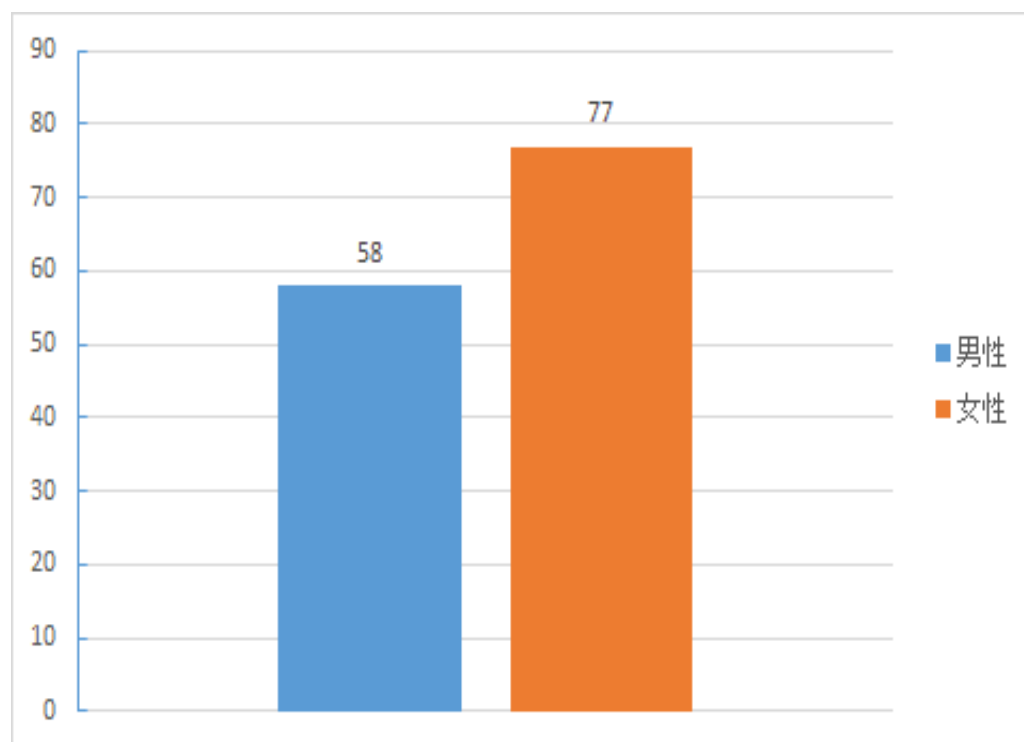


图 2 总样本用户男女数量显示图

### 5.2 所在学科分析

外语 48 人，历史 33 人，文学 22 人，哲学 11 人，图情 11 人，法学 5 人，政管 1 人，物理 1 人，社会学 1 人，建筑 1 人，电子科学与工程 1 人。

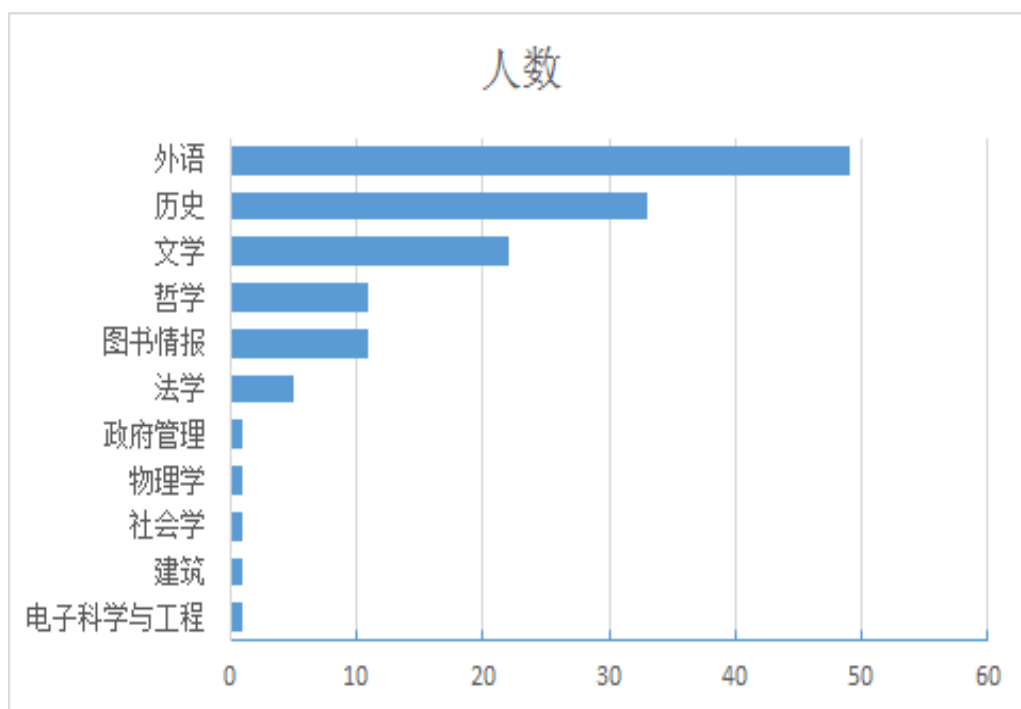


图 3 样本用户所在学科分布图

### 5.3 学位和职称分布分析

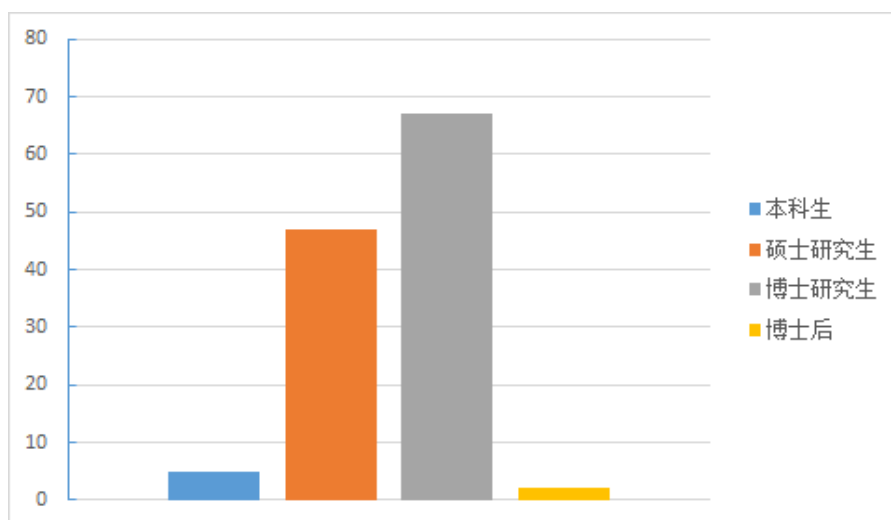


图 4 样本用户学位分布图

学生：121 人，其中本科生 4 人，硕士 48 人，博士 68 人，博士后 1 人

教职工：14 人，其中高级职称 5 人，中级职称 6 人，初级职称 3 人。



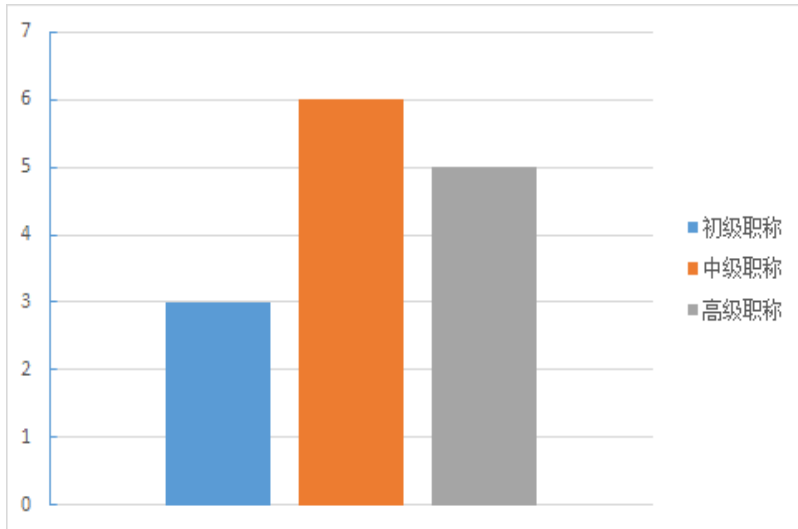


图 5 样本用户职称分布图

## 6 用户申请的定量分析

### 6.1 性别申请量分析

共收集到用户申请总量为 1629 条，其中男性申请量 660 条，人均 11 次；女性申请量 969 条，人均 12 次。

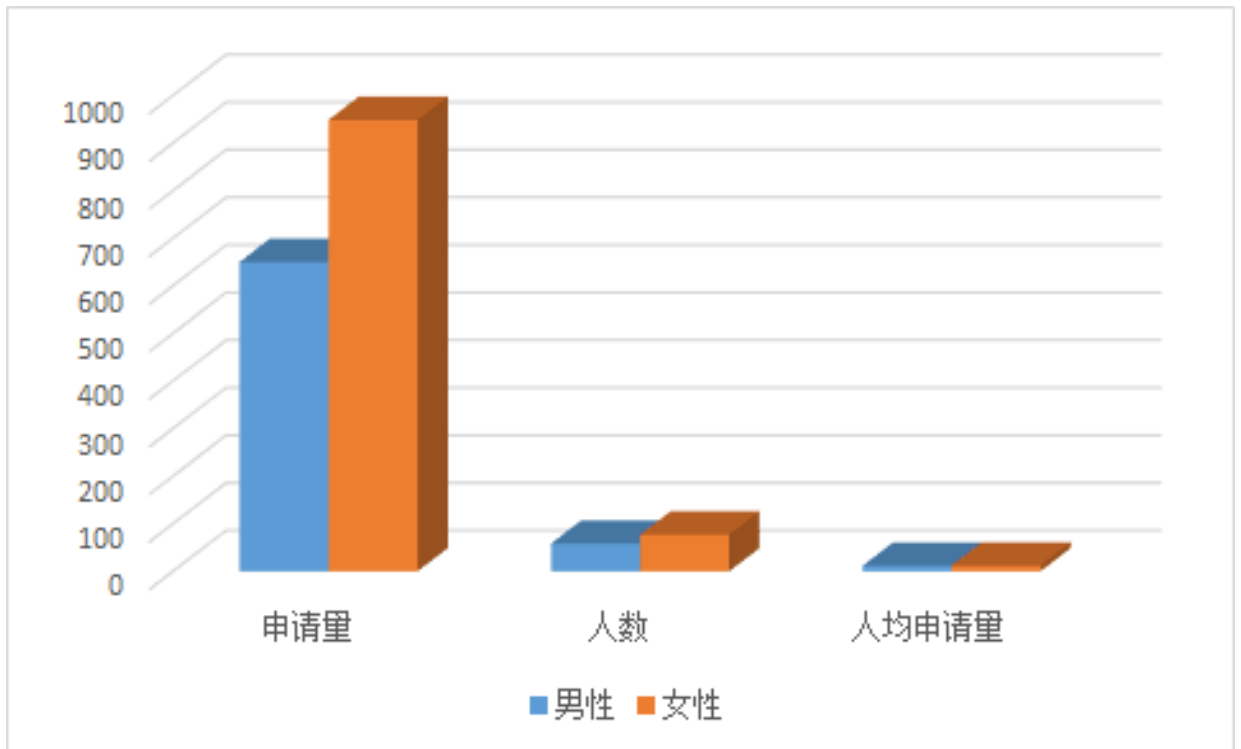


图 6 用户申请的一般特征

### 6.2 申请的学科特征分析

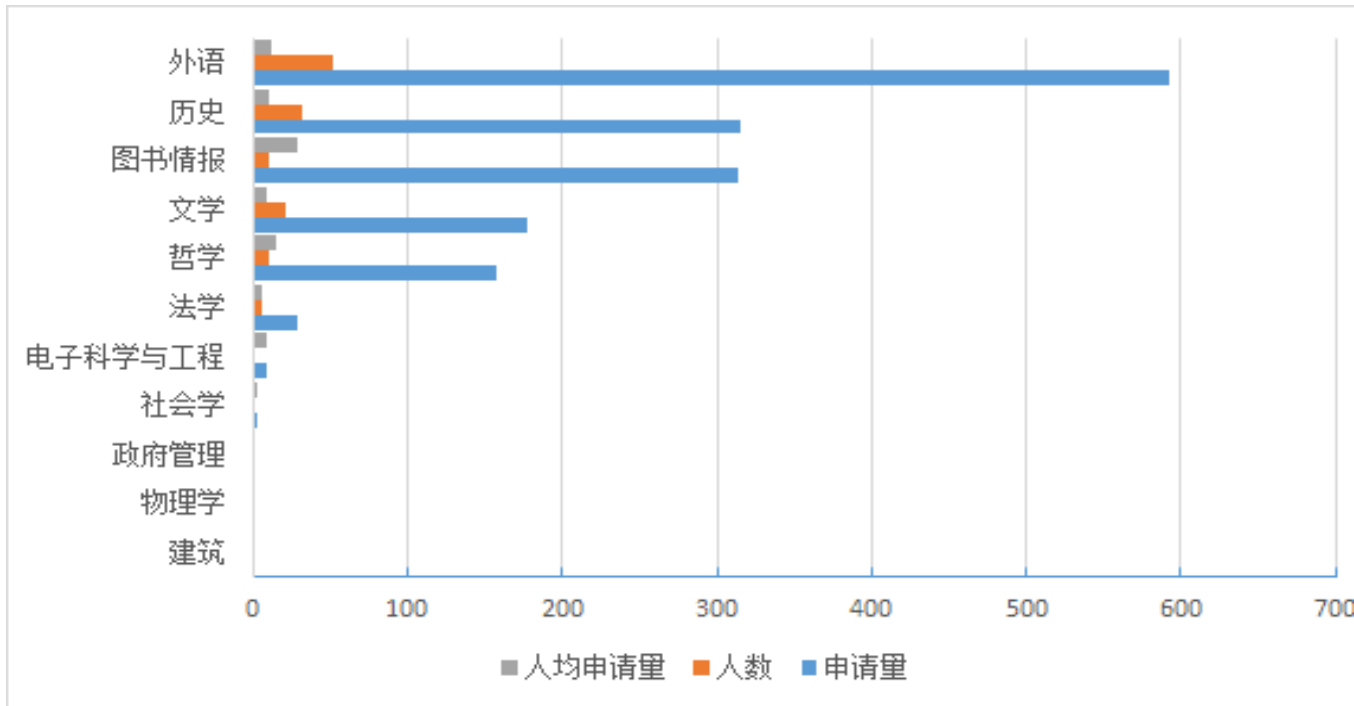


图 7 用户申请的学科分布图

外语 593 条，历史 515 条，图书情报 313 条，文学 178 条，哲学 158 条。

### 6.3 申请的学位和职称分析

本科生 39 条，硕士 334 条，博士 1028 条，博士后 9 条。其中博士人均申请最高，为 15 条。

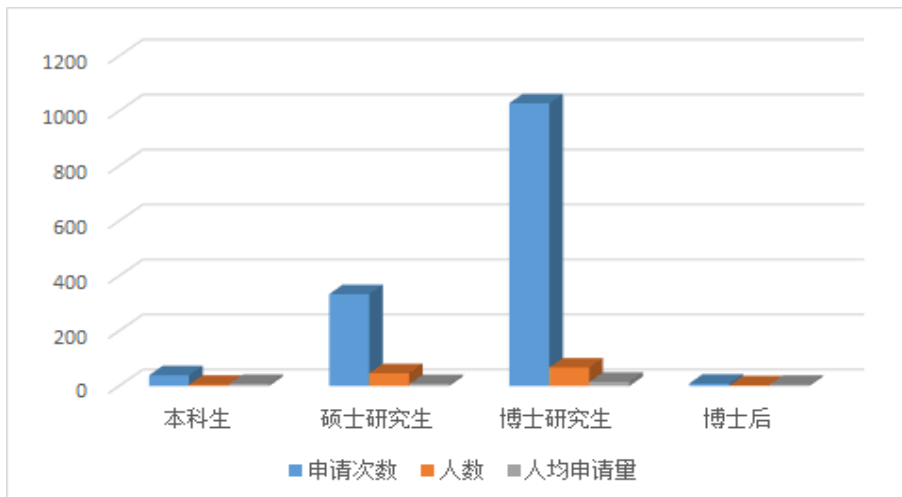


图 8 用户申请的学位分布图

申请的职称分析：初级 95 条，中级 76 条，高级 18 条，其中初级职称人均申请量最高，为 31 条。

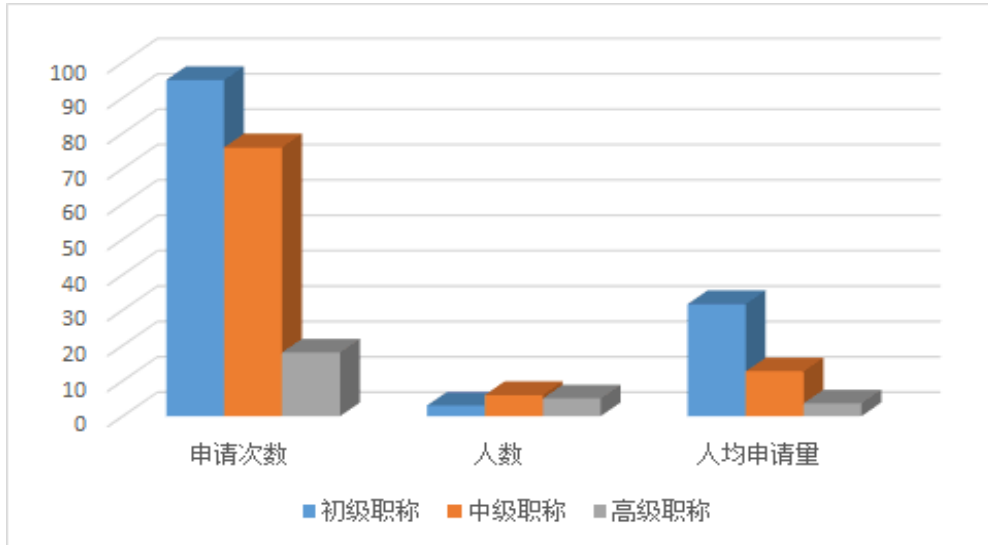


图9 用户申请的职称分布图

#### 6.4 用户申请量和申请次数

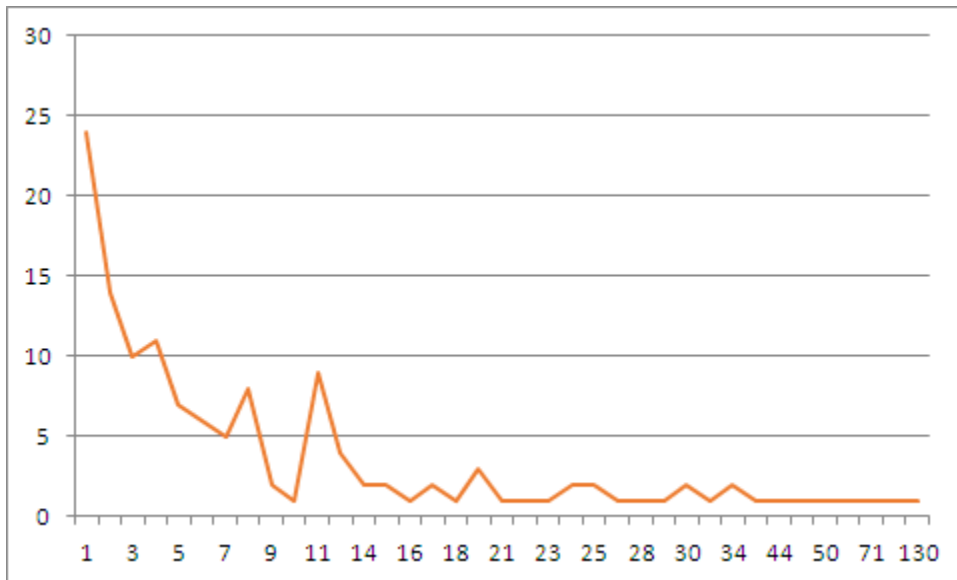


图10 用户申请量和申请次数图

有个人申请量最高达到130次，有24个人申请为1次，人均申请量是11.8次。

### 7 有科研产出(硕博士论文或科研发文)和无科研产出的用户和申请

#### 特征分析

本研究把有硕博士论文定题或有科研论文发表的用户定义为有科研产出的用户，其余为无科研产出用户。

#### 7.1 有科研产出用户的一般特征

##### 7.1.1 性别申请分布

男性34人，申请量450条，女性28人，申请量456条。

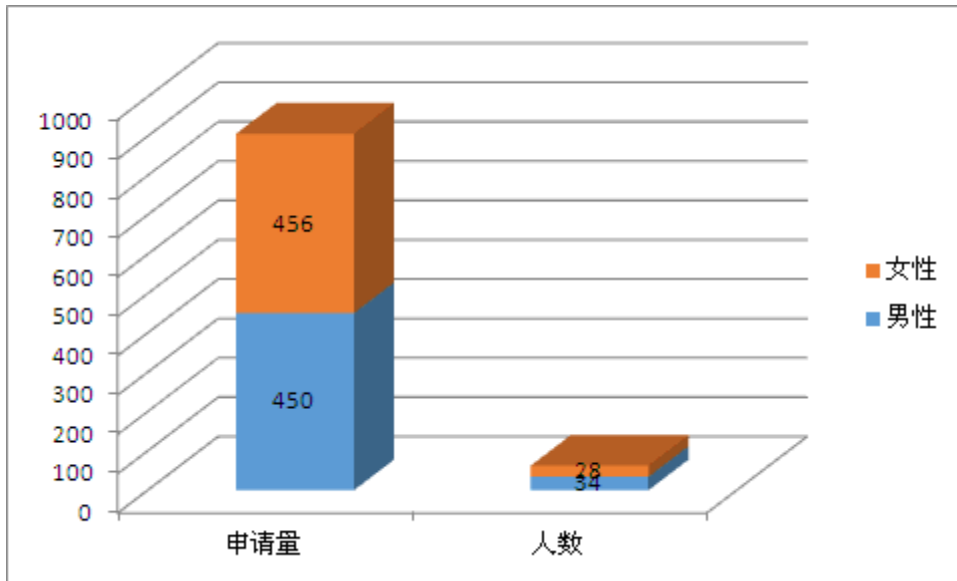


图 11 有科研产出用户的一般特征

7.1.2 有科研产出用户申请的学科分布

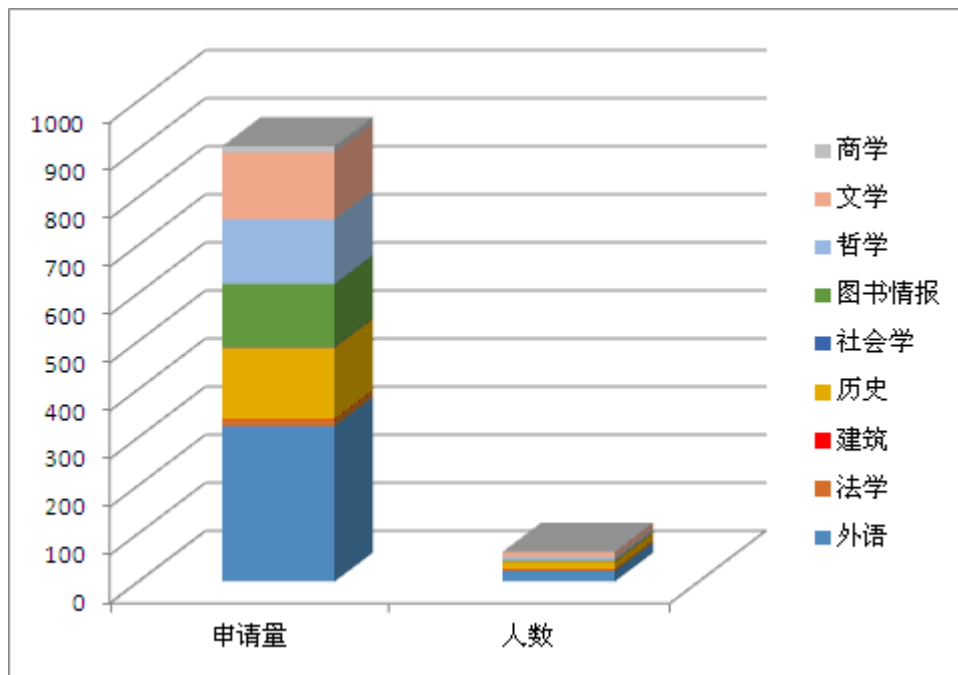


图 12 有科研产出用户学科分布

7.1.3 有科研产出用户申请的职称和学位分布

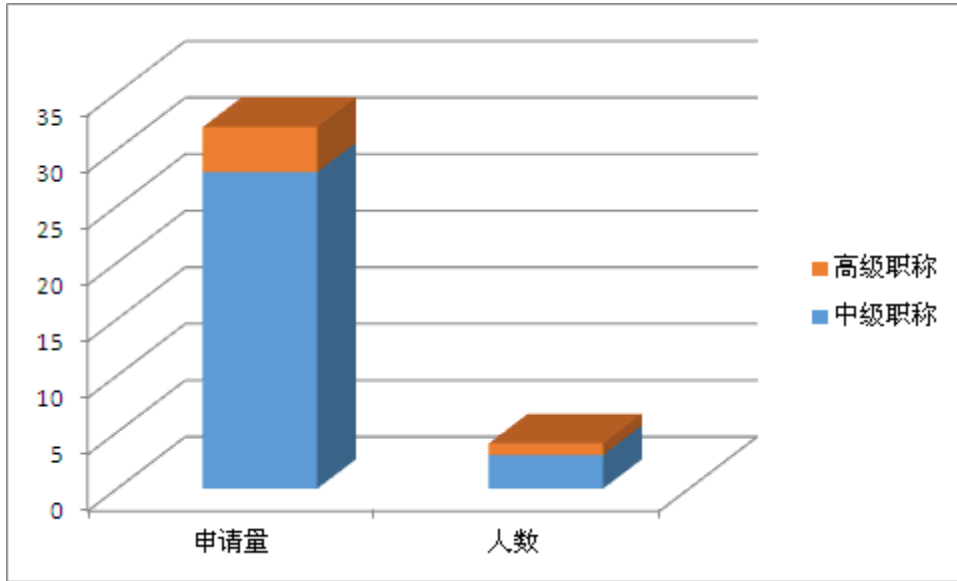


图 13 有科研产出的职称分布

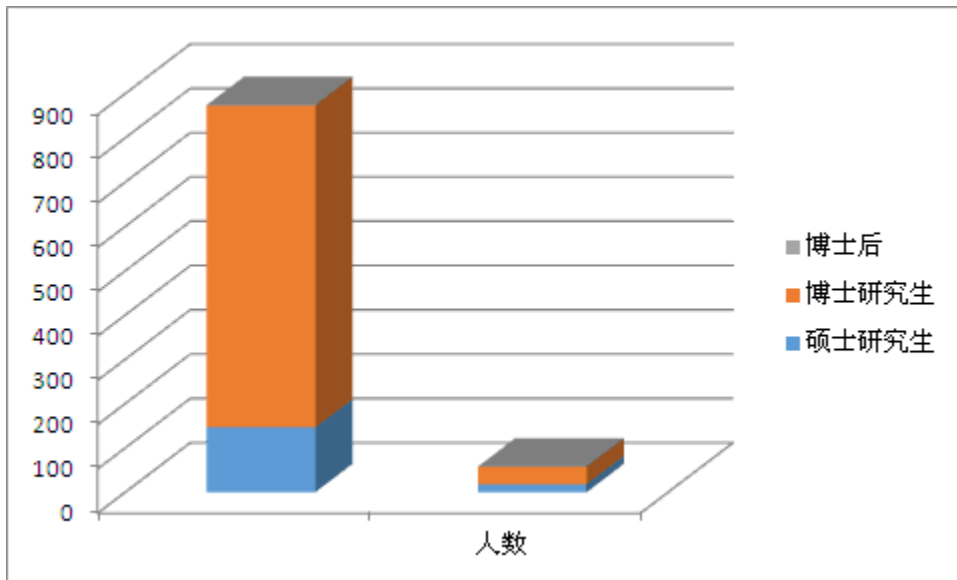


图 14 有科研产出的学位分布

## 7.2 无科研产出的用户的一般特征

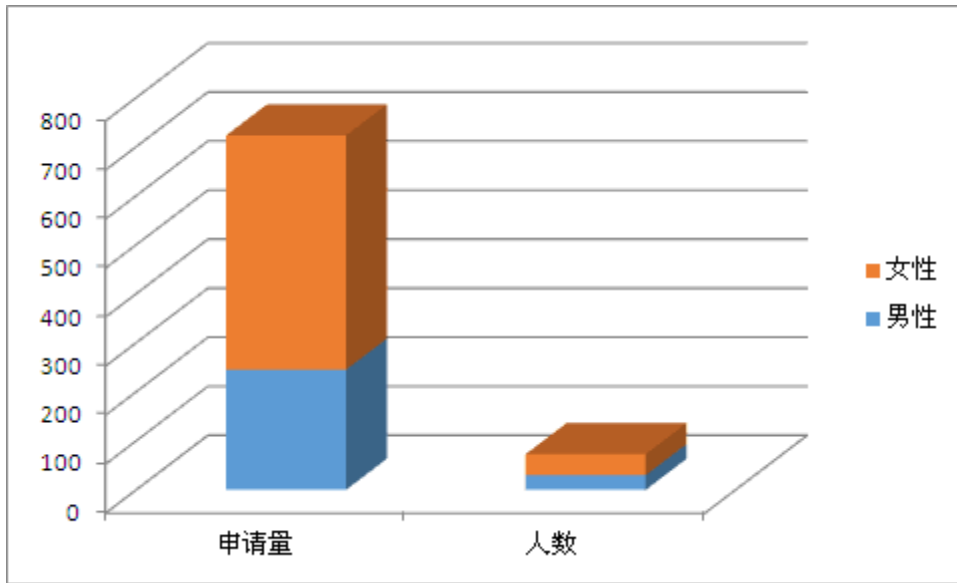
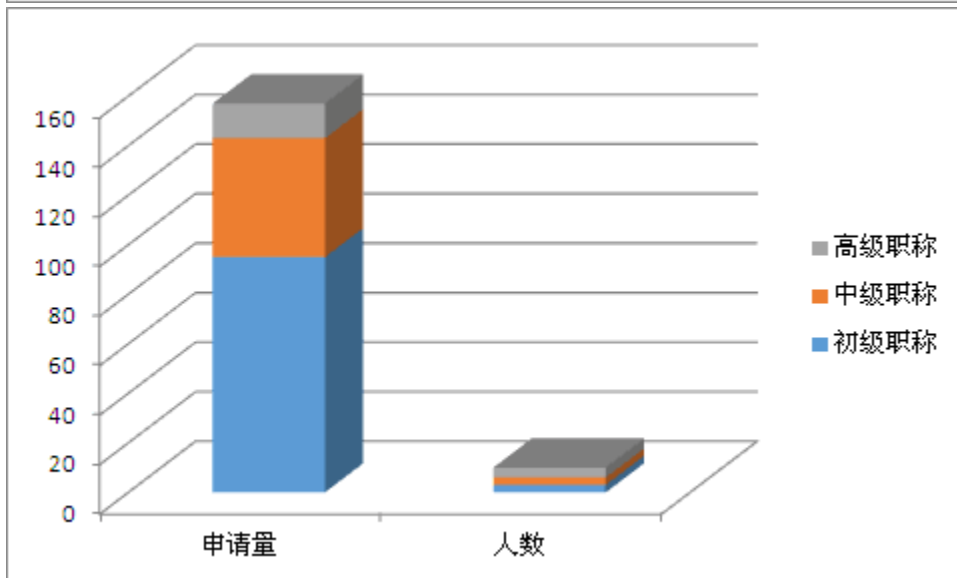
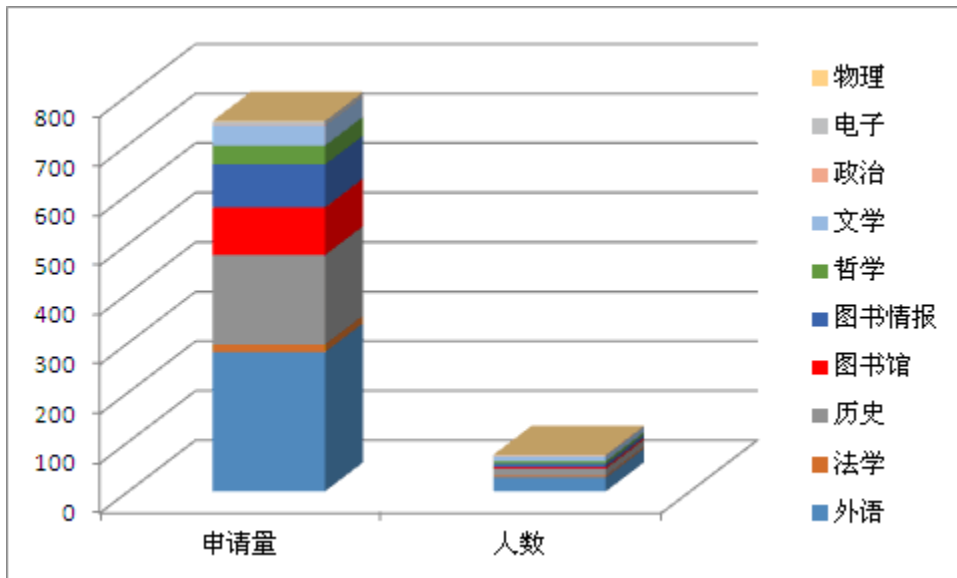


图 15 无科研产出的一般特征



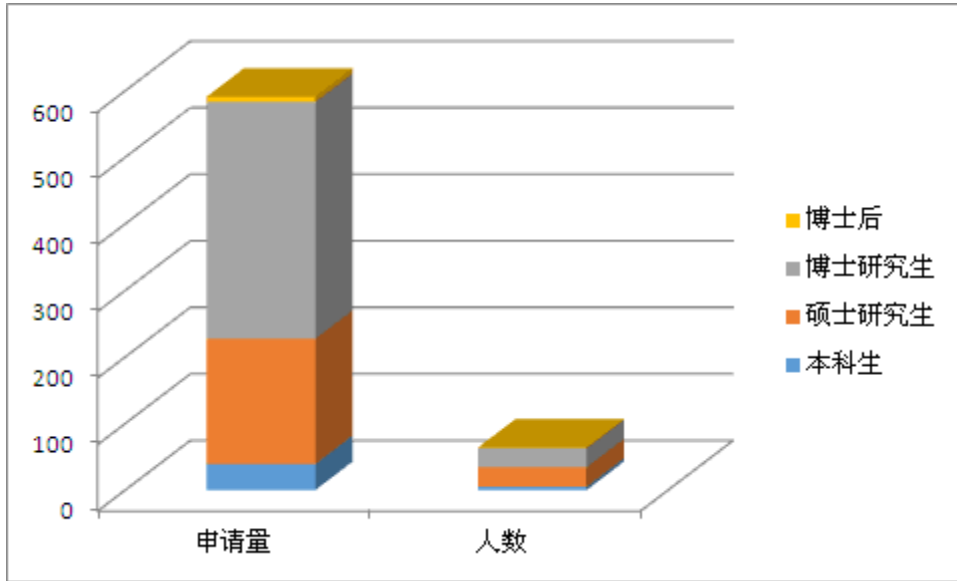


图 16 无科研产出的学科，职称，学位分布图

### 7.3 有产出和无产出的申请量和用户对比

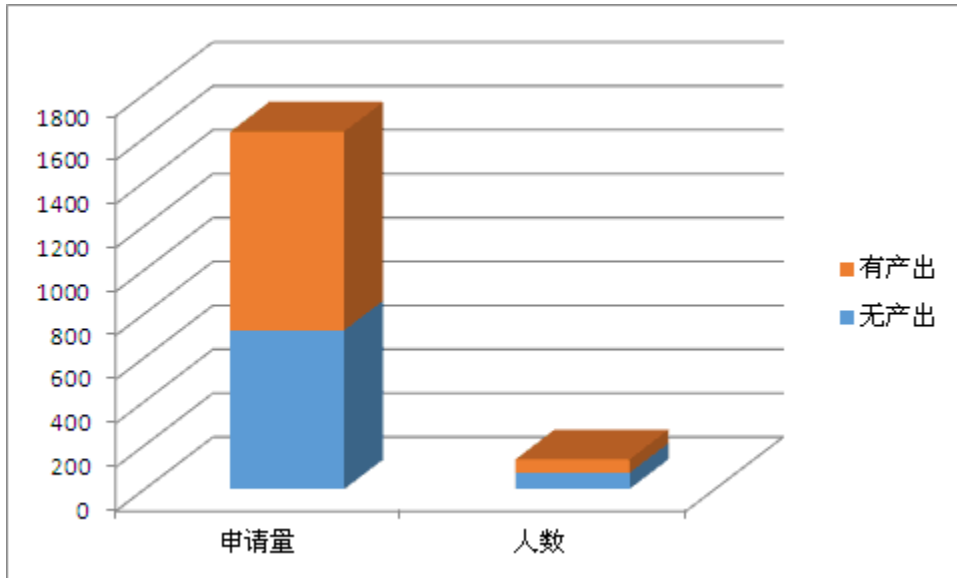


图 17 有产出和无产出的申请对比图

## 8 CASHL 申请资源和用户研究方向之间的文本相关度

研究者采用邮件对所有样本用户进行随访，获得了有科研产出用户提供的硕博士论文定题或发表论文题目，收集这些用户 CASHL 申请的图书与文献题目，然后采用 ictacas 和 standford 软件进行语义分词，采用 VosViewer 可视化分析工具绘制文本相似图谱，尝试做文本相似研究。如图所示





姓名	硕博士/论文题目	申请数	CASHL 图书	CASHL 文献
丁宇峰	专利质量的法律控制研究	书 3 条 文献 0	1)Patent ethics :litigation /David Hricik 2)The patent crisis and how the courts can solve it /Dan L. Burk and Mark A. Lemley. 3)Patent law and policy :cases and materials /Robert Patrick Merges, John Fitzgerald Duffy.	
沈浩	民事诉讼中科学技术的事实认定	书 11 条 文献 0		

图 20 研究方向数据来源截图

## 9 用户申请的时间变化特征

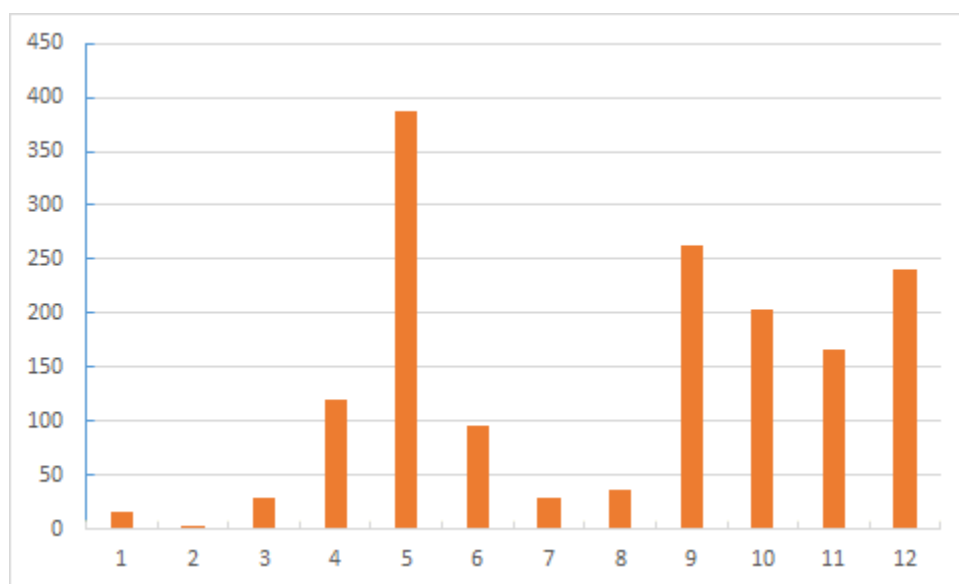


图 21 用户申请的时间变化趋势图

用户的申请高峰出现在 5 月份，为 388 条；其次为 9 月份，262 条；12 月份的 240 条；10 月份的 204 条和 11 月份的 167 条。

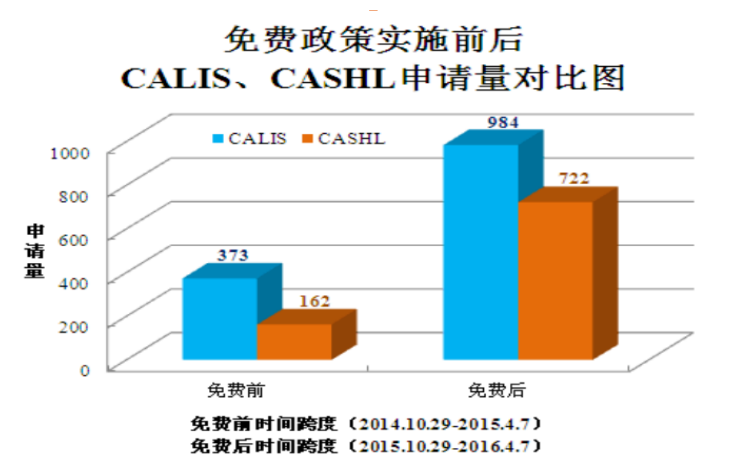


图 22 免费政策实施前后申请对比图

## 10 用户申请量、花费和科研产出的关系

为了分析申请量、性别、学位职称、学科、申请类型和申请次数对用户科研产出是否存在影响或存在相关性，采用 IBM SPSS20.0 软件作独立样本 t 检验和 Spearman 相关分析，结果显示：两组(有科研产出组 vs. 无科研产出组)的申请量为  $14.6 \pm 22.9$  vs.  $9.7 \pm 11.1$ ,  $p < 0.05$ ；无科研产出组中(女性 vs. 男性)的申请量为  $11.4 \pm 13.4$  vs.  $7.6 \pm 1.4$ ,  $p < 0.05$ ；其余影响因素如学位职称、学科和申请类型，因为涉及的一些样本人数  $< 10$  个，样本数小，影响统计意义，因此不作计算。

在所有申请中，无科研产出的申请量为 723 条，占到 44.38%，人均申请 9.9 条，共花费 7464.5 元。

## 11 被拒绝的申请分析

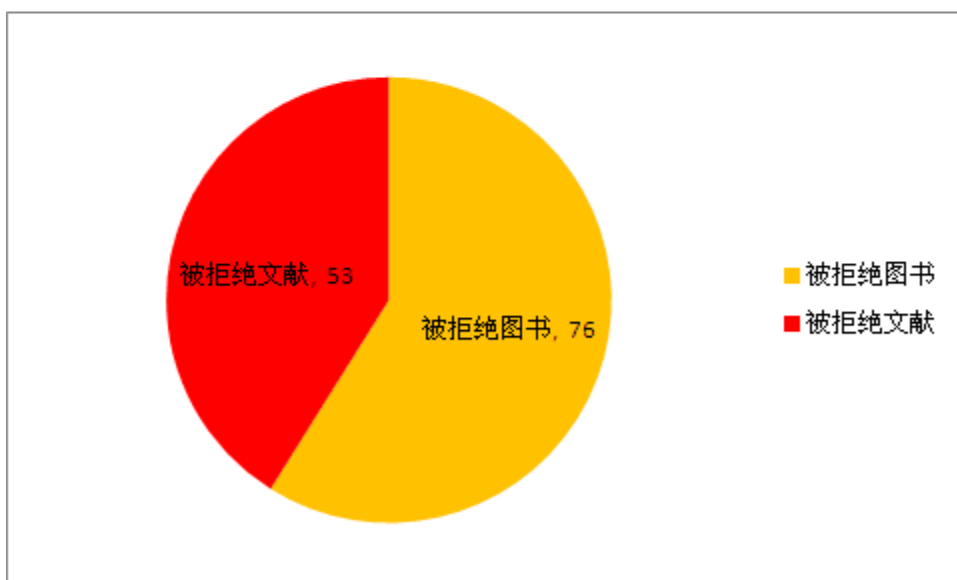
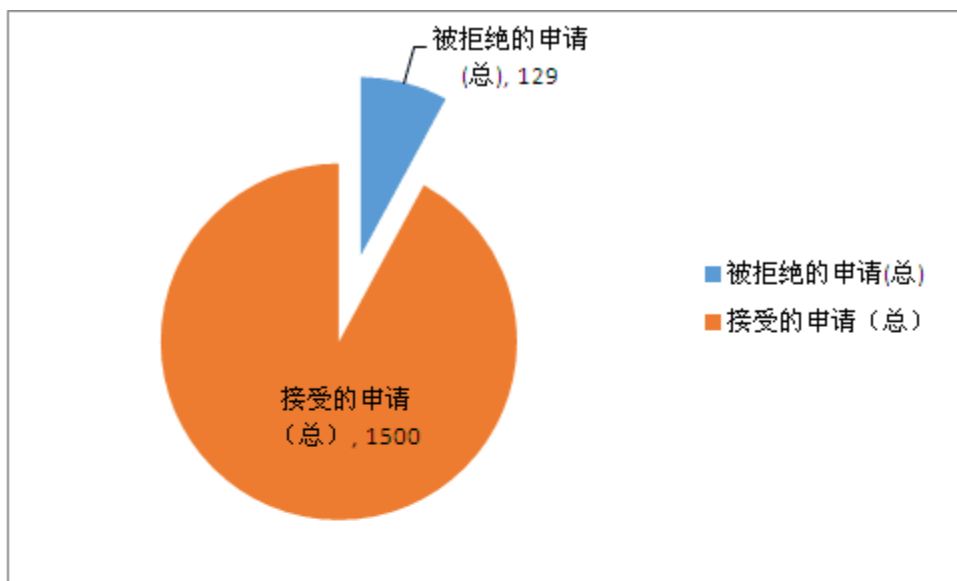


图 23 被拒绝申请的比例和文献类型显示图

所有样本用户共申请 1629 条，其中 129 条申请被拒绝，包括 76 条图书申请和 53 条文献申请。在被拒绝的申请中，因为馆藏原因（无馆藏、丢失、缺乏、）有 53 条；因为图书原因（出借/使用中、不流通、在装订、未上架、缺乏版权许可、状况不好）有 59 条；因为系统原因（包已发送成功）10 条；因为同学提供信息问题 5 条；其他原因 2 条。

拒绝原因	无馆藏	出借/使用	不流通	丢失	包已经发送成功,但是对方的状态不能接收这个	缺乏	未查指定文献	在装订	其他	未上架	缺乏版权许可	状况不好
------	-----	-------	-----	----	-----------------------	----	--------	-----	----	-----	--------	------

					包							
数量	34	22	23	10	10	9	5	5	2	5	2	2

表 1 被拒绝申请的各种原因

## 12 讨论

研究结果显示南京大学 CASHL 资源申请的所有用户中，硕博士学位的用户对 CASHL 资源的需求较大，其中博士研究生对 CASHL 资源的申请次数和使用频率最高。学科排名中，外语，历史，文学位列前三甲；职称分布中，中级职称用户的申请量最高，因此根据本研究样本，南京大学的用户群主要以博士生和中级职称用户为主。

其次根据用户提供的科研产出和硕博士定题信息，将样本人群分为有科研产出组和无科研产出组，通过比较，有科研产出组的申请量和申请次数明显高于无科研产出组，差异显著，有统计学意义( $p < 0.05$ )，同时通过 Spearman 相关性检验，显示有无科研产出和申请量密切相关，相关系数为 0.98。在无科研产出组中，女性的申请量和申请次数明显高于男性的申请量和申请次数，差异显著( $p < 0.05$ )。

本研究通过文本挖掘和可视化工具分析有科研产出用户的论文定题和文献申请，研究结果显示论文定题与申请文献之间有超过 80% 的文本相似度，说明 CASHL 西文资源能够很好地满足南京大学硕博学位论文的研究需求，对研究型用户提供了非常好的资源获取平台。其中一些硕博论文的定题研究方向新颖，但是本馆的馆藏不能满足这些研究者的资源需求，在利用 CASHL 资源满足他们研究需求的同时，我们也将这些用户所有的申请资料提交给采访部，以丰富和多元化馆藏，并且将这些研究者的研究方向和申请作为长期随访的研究对象。

根据申请量和申请时间结果显示，我校 CASHL 资源的申请量主要集中在 5 月份，9-12 月份，其中 5 月份是本校硕博学位论文预答辩和答辩月，很多硕博用户在最后提交论文前，均需要再一次进行文献查新，补充新的研究文献。9 月份是开学月，同时也是 CASHL 教师节优惠月，因此出现申请量的高峰是合理的。然而，研究结果显示 3 月份南京大学的申请量并不显著，甚至和寒暑假的申请量差不多。3 月份是南京大学硕博学生开题月，是申请项目报告的提交月，同时也是 CASHL 优惠活动月，这项结果显示南京大学馆际互借组在未来的工作中，需要大力加强每年 3 月份的推宣力度，扩大 CASHL 的学术影响力。

从图书馆来说，为了最大程度地满足我校研究者对资源获取的需求，2015 年 10 月，我馆推出了馆际互借费用全部免费的政策。该政策施行后，CALIS 和 CASHL 的申请量直线上升，与 2014 年同时间段相比，2015 年的申请量是 2014 年的 3-4 倍。因此 CASHL 申请量增加，用户对服务需求的增加，图书馆的政策和费用是非常重要的关键因素。

目前开世览文的文献服务还属于资源提供服务，就我馆来看，丰富而且免费的服务和资源对于处于初级和中级阶段的科研用户，具有较强的吸引力。然而，根据学位和职称分布显示，我馆高级职称的用户人数并不多，本文分析可能存在以下原因：

1. 一些高级职称用户自己不注册，委派学生注册并且提交申请；

2. 由于馆员的知识层次和简单的资源提供方式,不能给高级研究用户带来更多创新性型的研究思维驱动,因此暂时不能吸引到更多的高级职称人群;

3. 由于存在专业知识壁垒,在高级研究人员中开展 CASHL 推宣同时展开互动,对图书馆员来说有一定难度。

4. 由于样本数量不够大,此研究结果不一定能代表整体情况。

一直以来,如何针对研究型人群,包括硕博士用户和高级研究人员,进行学科服务和有效的资源提供和补充,是图书馆一直积极思考和亟待解决的问题。本研究认为应该积极组织参考咨询的学科服务组和馆际互借组联手拓展,从科研服务和有效资源双重角色入手,以此不断满足高层次人才科研需求。

在用户统计和分析过程中,还发现有 73 位用户没有任何科研产出,但他们的申请量为 723 条,占到了 44.38%,而且花费了 7464.5 元。这些用户一直在使用 CASHL 资源和服务,通过深度访谈和邮件随访,我们了解到一些用户因为刚刚入学,处于查找文献的初级研究阶段,一些用户主要是帮别人查文献,但是还有一些用户对自己的研究方向和定题是茫然的,而且他们所占的比例并不低(25%)。面对这些在研究方法或者研究领域没有思路的用户,图书馆的学科服务馆员,馆际互借组的文献提供馆员和院系的导师,可以形成一个三方互动的团队,从研究思路(导师)-资料查找和提供(馆际互借员)-资料阅读和研究方法(学科馆员),对该类用户开展一对一的个性化辅导,从而不断提升图书馆的学科服务水平。

根据 CASHL 申请中 129 条被拒申请的原因分析,本研究认为 CASHL 馆藏资源在数据揭示和平台管理上,还有一些方面亟待完善。首先应积极协调各 CASHL 参与馆的可借阅资源,从图书馆自身出发,完善各图书馆提交的 CASHL 馆藏揭示,定期做好馆藏更新维护。

我馆目前已经召开了编目部,采访部,和馆际互借数据维护组三方的讨论会,出台了三方协同管理规划方案,从采编数据开始,对南京大学所有的 CASHL 馆藏资源进行修改和完善,优化馆藏资源揭示。其次应积极改进 CASHL 平台的图书检索功能,分阶梯开放各馆可获取图书的额度;开放更多古籍,民国文献或者各馆的特藏优势资源。

作为最早加入开世览文体系的华东北区域中心,南京大学图书馆自 2004 年以来一直致力于完善馆际互借服务质量,图书馆领导非常重视读者的满意度。在 2015 年,馆际互借小组进行重组,馆际互借免费政策重磅出台,通过对该项目的实施,南京大学图书馆希望能够对服务用户拥有更深的认识,尝试把研究成果用于平时的工作中,更好地提升服务工作的品质。

然而由于本研究团队的研究经验和时间不足,本研究的样本量不够大,研究周期不够长,在今后的研究中本研究团队会努力克服,将研究结果不断完善。

## 13 结论

通过采用实证方法和对客观数据的分析与讨论,可以得到以下研究结论:

1. 南京大学的用户群主要以博士生和中级职称用户为主;
2. 有无科研产出和 CASHL 的申请量密切相关,其中有科研产出的用户申请量明显高于无科研产出用户;
3. CASHL 西文资源能够很好地满足南京大学硕博学位论文的研究需求,对研究型用户提供了非常好的资源获取平台。
4. 一些硕博学位论文的定题研究以及 CASHL 资源的申请目录有助于丰富和多元化本馆的馆藏建设。
5. 在未来的推宣工作中,馆际互借组需要大力加强 3 月份的推宣力度,扩大 CASHL 的学术影响力。
6. 图书馆的政策和费用是促进 CASHL 申请量增加,用户对服务需求增加的重要因素。

7. 尽可能地积极组织参考咨询的学科服务组和馆际互借组联手拓展，从科研服务和有效资源双重角色入手，以此不断满足高层次人才的科研需求。
8. 面对没有任何科研产出的用户，面对在研究方法或者研究领域没有任何思路的用户，图书馆的学科服务馆员，馆际互借组的文献提供馆员和院系的导师，可以形成一个三方互动的团队，从研究思路(导师)-资料查找和提供(馆际互借员)-资料阅读和研究方法(学科馆员)，对该类用户开展一对一的个性化辅导，从而不断提升图书馆的学科服务水平。
9. CASHL 馆藏资源在数据揭示和平台管理上，还有一些方面亟待完善。首先应积极协调各 CASHL 参与馆的可借阅资源，从图书馆自身出发，完善各图书馆提交的 CASHL 馆藏揭示，定期做好馆藏更新维护。其次应积极改进 CASHL 平台的图书检索功能，分阶梯开放各馆可获取图书的额度；开放更多古籍，民国文献或者各馆的特藏优势资源。

## 14 致谢

在数据采集集中，感谢两位学生助理汪辉和王宜鸿同学。在 Vosviewer 绘图中，感谢信管学院的研究生威尔鹏同学编程并绘图。

## 参考文献

- [1]邱均平,马凤,. 中国高校在建设世界一流大学过程中的进步和问题——基于 2011 年《世界一流大学与科研机构学科竞争力评价》的分析[J]. 中国高教研究,2012,(1):17-22.
- [2]<http://www.cashl.edu.cn/portal/html/article19.html>
- [3]王浩,. 图书馆学学术规范研究综述[J]. 图书馆建设,2014,(1):5-9.
- [4]卢滨玲. 运用调查法建立图书馆读者信息反馈机制[J]. 图书情报知识,1995,(1):38-39.
- [5]<http://www.leydesdorff.net/software.htm>

## 附录

---

---

### 表格目录

---

表 1 被拒绝申请的各种原因.....	20
---------------------	----

### 图表目录

---

图 1 研究设计路线图.....	6
图 2 总样本用户男女数量显示图.....	7
图 3 样本用户所在学科分布图.....	8
图 4 样本用户学位分布图.....	8
图 5 样本用户职称分布图.....	9
图 6 用户申请的一般特征.....	9
图 7 用户申请的学科分布图.....	10
图 8 用户申请的学位分布图.....	10
图 9 用户申请的职称分布图.....	11
图 10 用户申请量和申请次数图.....	11
图 11 有科研产出用户的一般特征图.....	12
图 12 有科研产出用户学科分布图.....	12
图 13 有科研产出的职称分布图.....	13
图 14 有科研产出的学位分布图.....	13
图 15 无科研产出的一般特征图.....	14
图 16 无科研产出的学科、职称、学位分布图.....	15
图 17 有产出和无产出的申请对比图.....	15

图 18 硕博士论文研究定题可视化图.....	16
图 19 CASHL 申请图书或文献的语义可视化图 .....	16
图 20 研究方向数据来源截图 .....	17
图 21 用户申请的时间变化趋势图 .....	17
图 22 免费政策实施前后申请对比图.....	18
图 23 被拒绝申请的比例和文献类型显示图 .....	19