

项目结题验收单

1 专家验收表（主持人所在单位组织 3-5 名专家对项目进行验收、自评。）

项目名称	用户访问电子资源行为数据实证研究—以内蒙古农业大学为例				
主持人	孙利芳	职务/职称	副研究馆员		
所在单位	(加盖公章) 内蒙古农业大学图书馆				
专 家 意 见	<p style="text-align: center;">  </p> <p>该项目组搭建了电子资源分析平台，以通过平台获取到的内蒙古农业大学读者访问电子数据库行为数据为基础，从读者的访问量、访问的热门学科、访问的学级、访问排行榜等多个维度进行对比分析形成分析报告，图书馆可根据使用统计分析报告了解电子资源的访问情况、读者的使用行为，为电子资源的合理使用提供参考数据。</p> <p>项目小组成员通过近一年的时间完成了具体工作，对电子资源使用数据进行了分析，并为馆里提供了多份月度、季度、年度统计分析报告。</p> <p>项目圆满完成了既定目标，专家组一致同意项目通过验收。</p> <p style="text-align: right;">(如需要可增加页数)</p>				
专家签字	包峰	旭荣苑	郭冬梅	王小丽	杨毅
职务/职称	研究员	研究员	副研究馆员	副研究馆员	副研究馆员



项目编号：2023039

CALIS 全国农学文献信息中心研究项目 结题报告

项目名称：用户访问电子资源行为数据实证研究—以内蒙古农业大学为例

项目关键词：用户行为；电子资源；数据分析；智慧图书馆

项目单位(盖章)：内蒙古农业大学图书馆

通信地址：(详细地址含邮编) 内蒙古呼和浩特市昭乌达路 306 号内蒙古农业大学图书馆

项目主持人：孙利芳

联系电话：13674825220

电子邮件：308587114@qq.com

提交日期：2024 年 4 月 25 日

(结题报告含有以下 5 部分内容，其他内容根据项目情况可增加,字数不少于 4000 字)

题目：用户访问电子资源行为数据实证研究—以内蒙古农业大学为例

关键词：用户行为；电子资源；数据分析；智慧图书馆

1 研究背景、目的及意义

随着“智慧图书馆”的出现，国内外图书馆界都开始重点关注自身的发展怎么和智慧一词相结合，智慧图书馆评估的对象包含了图书馆软硬件等物的智慧，还包含馆员、用户等人的智慧。对于基础设施和工具的智慧评估，情境感知、用户行为分析、数据挖掘等维度是具体体现，因此做好用户行为的记录与分析在智慧图书馆管理和服务中显得尤为重要。

互联网技术的发展以及大数据时代的到来，泛在化学习、科研环境下，电子资源与传统的纸本文献相比，更受广大师生青睐，已成为高校图书馆用户获取信息的主要来源。为更好地满足读者需求，推进一流学科建设，内蒙古农业大学图书馆对电子资源投入的购置费在文献资源购置总经费中占比较高，占文献资源购置费 65%以上，基本上呈现逐年攀升的趋势。因此电子资源用户行为使用统计就是图书馆比较受关注的一个主题。

访问电子资源的行为数据能精准地反映出用户在教学、科研等方面的信息需求，现阶段，绝大多数资源数据库使用统计是依赖数据商提供，

由于数据商无法获取到读者的具体信息,准确性及数据分析维度都没有达到图书馆服务学科的需求,从而数据库商提供的使用统计报告无法分析到读者的行为数据。而且数据库商提供的报告中,指标数值的准确性、真实性无法保障。那么建立一条自有的电子资源分析的有效途径,有利于电子资源绩效评估工作的规范化、有利于电子资源使用的针对性、有利于智慧图书馆的管理和服务。

本课题利用数据分析方法对获取到的读者行为数据按照确定的电子资源分析指标进行数据分析,可按月或年提供电子资源使用分析报告。图书馆可根据使用统计分析报告在资源的试用、采购阶段结合本校的学科发展以及读者与专家的推荐意见,综合考虑订购电子资源。在资源的访问、使用监护阶段,可以为图书馆随时提供电子资源使用的情况,了解读者的使用行为,对于使用量较低的电子资源,图书馆可考虑通过调研、宣传、培训的方式引导读者使用,及时发现问题,以保证资源的合理使用。

2 研究内容及方法（思路、方法、具体内容）

2.1 系统搭建

本项研究是基于软件代理的行为数据获取模式,基于地址重构的行为数据获取模式采用统一身份认证+地址重构的方式获取行为数据。图书馆在本地架设二台地址重构服务器,用户访问远程数据库时先进行学校统一身份认证,访问请求通过地址重构服务器中转后再转发给远程数据库;数据库返回结果也通过地址重构服务器中转重构后再返回给用

户。地址重构服务器记录所有用户行为，这样图书馆对服务器日志过滤解析获得用户行为数据。经过两个月的搭建与调试、测试，系统进入正常使用阶段。

为了保障网关系统、统计系统的稳定性，共搭建 2 台高性能服务器，其中一台作为域名分析平台服务器(16 核 32G 内存)，一台作为数据库分析服务器(8 核 16G 内存)。

搭建的平台运行界面如下：



2.2 指标确定

笔者查阅了相关文献，发现目前并没有用户数字资源使用指标的统一标准，经过课题组成员商讨初步定了从资源使用量、用户使用情况和资源内容三个方面对使用的整体状况进行了研究，确定了图书馆电子资源统计指标。

2.2.1 资源使用量指标：包括检索量、访问量等全局性使用量以及下载量指标。其中访问量的多少是反映用户利用电子资源意愿的直接指标。

2.2.2 用户使用情况指标：按使用资源的不同用户组进行统计，包括读者类型（教职工、研究生、本科生）、读者学院、读者学科专业、入学年份等指标，有了这些用户的基本信息，可以对用户进行不同纬度的划分，成为用户行为分析的重要指标。

2.2.3 资源内容指标：包括读者访问的期刊、访问的电子资源名称、检索词等。其中检索词的获取可以发现读者研究热点，可针对性开展培训等活动。

确定了指标后的读者行为查询界面如下：

序号	时间	用户名	外网用户名	内网用户名	姓名	性别	专业	年级	读者类型	院系名称	IP	标题	文章关键词	作者	作者单位	资源类型
1	04-30 23:58:35			201		男			教职工	水利与土木建筑工程学院	10.1	限沙式抽沙船设计及防淤效果分析	限沙式抽沙船 孔隙率 风沙流 防冲 双室 数据模拟	张展 张少安 徐 建刚 杨力 张亚 宇	1. 兰州交通大学机电工程学院 2. 兰州交通大学轨道交通工程教育部级重点实验室	期刊
2	04-30 23:58:05			2014		男			教职工	水利与土木建筑工程学院	10.1	煤矿开采影响下岩溶地下水化学特征及氯化钙研究	地下水 水化学特征 离子比值 水文 地球化学作用 岩石风化	隋海 魏朝超 魏 征	河北地质矿产勘查开发局第一地质大队	期刊
3	04-30 23:46:22		1104	2020116		男	动物科学	2020	本科生	动物科学学院	10.1	油桐在畜禽生产中的应用研究		刘卫刚 李超廷	1. 河南农业大学牧医工程学院 2. 河南农业职业学院畜牧兽医研究所	期刊
4	04-30 23:45:51			2020116		男	动物科学	2020	本科生	动物科学学院	10.1	亚麻自行滚动播种器对土壤团聚体变化的影响		李伟强 李惠玲 孙海强 王树福 王海波 张琳 姜 琴 汪龙博 王秀 英 刘玉兰	武汉轻工大学动物营养与饲料科学湖北省重点实验室	会议论文
5	04-30 23:41:50			20201		女	应用化学	2020	本科生	理学院	10.1	TiO2纳米颗粒改性及其光催化降解性氧化亚砷的产率	二氧化钛 光催化降解 选择性催化 降解率 稳定性	邢兵	浙江理工大学	学位论文
6	04-30 23:40:50			2020		女	工程造价	2020	本科生	水利与土木建筑工程学院	10.21	高校建筑改造工程实施阶段造价控制研究	高校 建筑 改造工程 造价 控制	马娜	北京建筑大学	学位论文
7	04-30 23:36:54			2005		女			教职工	人文社会科学学院	10.225	农业R&D投入与农业全要素生产率——一个空间溢出视角的检验与证据	农业R&D投入 农业全要素生产率 空间溢出效应	陈琦 邓美荣	华中大学经济管理与法学学院	期刊
8	04-30 23:36:40			2005		女			教职工	人文社会科学学院	10.221	农业科技资源分布与主要农业生产	农业科技资源 农业TFP 农业TFP 农业TFP 农业TFP	杨伟青 王修梅	桂林理工大学商学院	期刊

2.3 统一身份认证对接

对接了基于校园网作身份认证的数据库权限访问后，可以获取读者的个人行为数据，包括学院、专业、年级、性别等。利用这个途径获得用户访问网络资源行为的数据，并根据确定的电子资源统计指标对数据进行分析，可构建用户画像精准服务模式，通过数据赋能实现服务增值。

2.4 数据分析

行为数据的挖掘、分析是一项艰巨的任务，对已购置的且读者使用率较高中国知网作为测试数据库进行挂载，研究根据电子资源分析系统中数据库的使用数据，数据中包括了资源访问的 Num(访问数量)、Name(电子资源名称)、Keyword(关键检索词)；而读者院系、学级、姓名、专业等其他分析维度通过获取数字化校园数据中台中的数据进行碰撞产生。经过近一年的使用，对读者利用数据量进行对比分析，根据使用数据，比较分析数据库整体使用情况、读者使用行为等，并结合数据库类型、学科信息、学科分布、院系等对资源使用情况，进行了用户行为分析等深层次数据挖掘与分析，研究结果可作为数据库重要性评价的辅助参考。

有了这些汇总就可以得出哪些学院或者哪些群体在使用数据库，以及具体使用情况，以便为学科提供精准服务。

研究将系统平台获取的数据从时间、学科、年级、学院等多维度进行统计分析，全面、直观地揭示了 2023 年 6 月至 2024 年 4 月的数据库访问情况。



图 2 行为分析层次图

3 行为数据分析

访问量是最为直观的衡量数据库使用率的指标，通过对中国知网数据库近一年来的访问量进行统计，从图 3 可以看出数据库读者访问高峰期是 3 月和 6 月。整体来看，这种趋势有一定的规律性，与高校师生科研及学习的周期密切相关^[1]，由于 8 月和 2 月是放假期间，读者很多使用了校外访问系统，这部分数据目前还无法获取，故使用量相对比较少，4 月份是研究生、本科生提交论文前的时间，故使用量相对较多，但是本分析报告中 6 月份这个数据量最大有些不太符合这个规律性，可能与本系统刚刚上线有关，有待与今年数据进行对比分析，查找原因。由于系统在 11 月份时出现了故障，有近两周时间没有收集到数据，故 11 月份的数据也是不准确的，但是对总体数据量分析影响不大。



图 3 2023 年 6 月—2024 年 4 月中国知网数据库访问量

在系统获取数据的 11 个月里共有访问量 100 多万人次，访问人数达到了 11000 多人，占比学校总人数的三分之一。

从学院分析可知（详细数据见附件）读者访问量最多的学院是经济管理学院、食品科学与工程学院、人文社会科学学院。在这三个学院里人文社会科学学院读者比例占比很小，但是访问量却是前三，说明数据库需求量高，以后图书馆可以多开展读者的信息检索能力培训和文献推送服务。

从读者类型分析可知（详细数据见附件）研究生读者访问量最高，占总访问量的 50%，本科生使用量占 32%，今后可以引导本科生多使用数据库，对研究生可以展开培训提升其数据库使用技巧，提升检索效率。

从入学年份方面分析可知（详细数据见附件），2022 级读者访问量最高。但是对本科生和研究生的数据进行筛选、分析后发现 2022 级研究生使用量最多，与其他年级相比数据量差距一倍。但是本科生使用量最多的是 2020 级，且与其他年级差距最多的有 10 倍之多，少的也有 4 倍左右。说明在这个阶段读者文献资源需求量大，图书馆可以有针对性的在这个阶段进行数据库及信息检索培训。

4 项目成果（发表的文章、开发的软件、取得的实践效果等）

经过约二个月的安装、调试以及安全策略的调整在校内搭建了电子资源分析系统软件，通过近一年的使用，笔者对海量数据进行筛选、清洗、分析、统计形成了多份月报表，季度报表和年度报表。为图书馆馆领导对数据库使用情况提供了一定的参考，后附 2023.6-2024.4 电子资源分析报告。

5 总结

通过数据的分析可以把握用户使用高峰期, 瞄准、扩大核心用户群体^[2], 为其提供精准的电子资源推送及服务。但是用户行为分析系统的构建, 需要通过持续的行为数据积累以达到对行为分析的准确性和有效性。故一年的数据量对于分析来说数据量还是缺乏一定的说服力, 需要以后继续做深入的研究, 挖掘出有利于资源使用、学科发展的数据。

因系统对接了学校统一身份认证, 服务器数据中含有大量涉及个人隐私的敏感数据。根据数据安全法及个人信息保护法要加强数据安全保护, 以防数据泄露对读者造成不必要的影响。

6 参考文献

- [1]万莉,孙德福,刘士义. 高校图书馆电子资源使用情况分析研究——以山西师范大学外文数据库为例 [J]. 甘肃科技, 2024, 40 (01): 102-105. DOI:10.20156/j.cnki.2097-2490.2024.01.021.
- [2]吴亚平,李晓东,艾春艳. 资源发现系统用户行为特征分析——以北京大学图书馆为例 [J]. 大学图书馆学报, 2023, 41 (01): 70-77. DOI:10.16603/j.issn1002-1027.2023.01.010.
- [3]李萍萍,葛敬军,谭明友. 读者行为大数据分析模型应用于高校图书馆创新服务研究——以济宁医学院图书馆为例 [J]. 图书馆工作与研究, 2022, (10): 61-66. DOI:10.16384/j.cnki.lwas.2022.10.016.
- [4]刁羽,薛红. 高校图书馆典型用户群体电子资源行为数据分析实证研究——基于创文图书馆电子资源综合管理与利用系统 [J]. 新世纪图书馆, 2022, (07): 59-64+71. DOI:10.16810/j.cnki.1672-514X.2022.07.010.
- [5]孙瑾. 电子资源绩效评估系统对比分析研究——以山东大学图书馆为例 [J]. 图书馆杂志, 2022, 41 (06): 83-87+115. DOI:10.13663/j.cnki.lj.2022.06.011.
-

[6]王大阜,邓志文,贾志勇. 新时期电子资源使用统计系统研究与应用 [J]. 图书馆论坛, 2023, 43 (01): 87-95.

[7]陈心怡. 基于大数据的高校图书馆智慧搜索服务模式研究[D]. 天津理工大学, 2022. DOI:10.27360/d.cnki.gtlgy.2022.000165.

[8]陈佳宇,郭建伟,燕娜,等. 数据库资源使用统计与用户检索行为分析——以超星知识发现系统为例 [J]. 创新科技, 2018, 18 (02): 79-82. DOI:10.19345/j.cxkj.1671-0037.2018.02.022.

附件 1 电子资源分析报告

内蒙古农业大学电子资源分析报告（2023. 6-2024. 4）

1 数据展示及分析

在大数据的背景下如何对读者访问电子资源的行为进行分析和统计，是图书馆做好服务工作、提高服务质量和读者满意度的前提和重要保障。通过采用基于私有云技术的软件代理行为数据获取模式，添加数据挖掘功能，对读者访问电子资源的行为进行统计并分析，为后续购买电子资源和提高读者的满意度提供数据依据。

此报告为用户在 2023 年 6 月 01 日至 2024 年 4 月 30 的电子资源行

为数据所做的分析统计，具体分析说明如下：

2 资源使用分析

从电子资源库的行为以及资源类型进行多角度、多维度的数据分析，通过对校内资源库使用的综合分析，可以为图书馆数字资源的建设和提高用户利用率提供必要的参考和指导。

图书馆电子资源分析平台搭建试运行后，首先挂接了中国知网数据库，数据比较完整，故以下分析仅以中国知网数据库为例进行分析。

2.1 使用量

表 1 电子资源使用详情表（2023.06-2024.04）

所属总库	总库行为总量	总库浏览次数	总库检索次数	总库下载次数
中国知网	1118096	505201	256073	240627

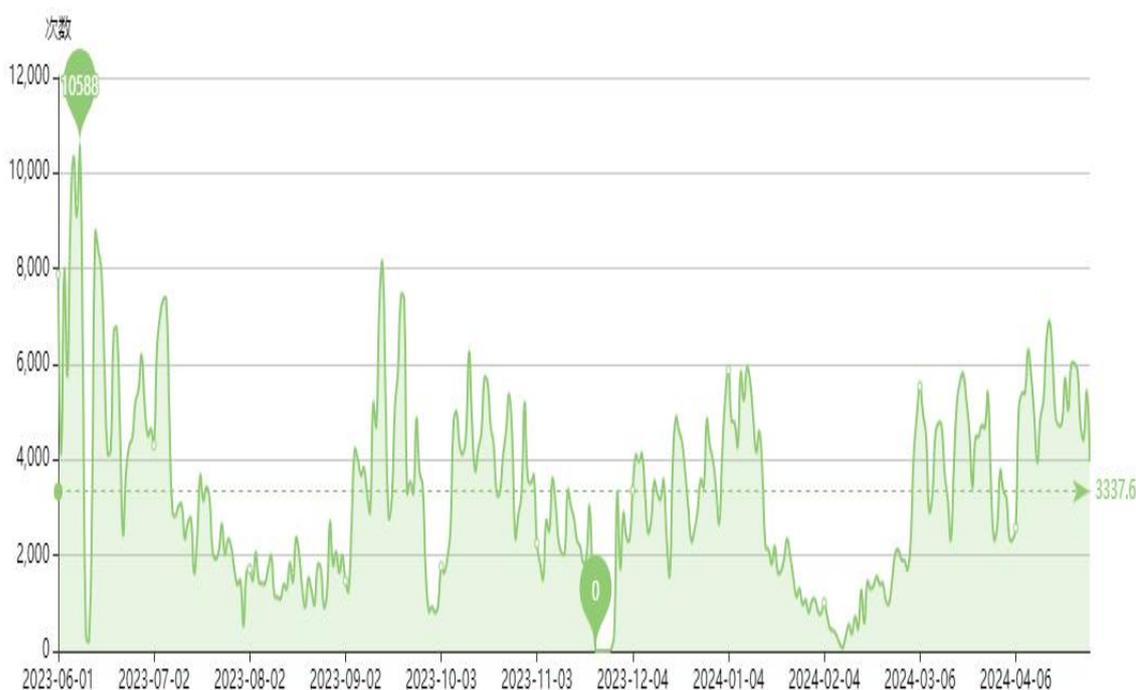


图 1 中国知网数据库用户访问量（2023.06-2024.04）

2.2 检索词

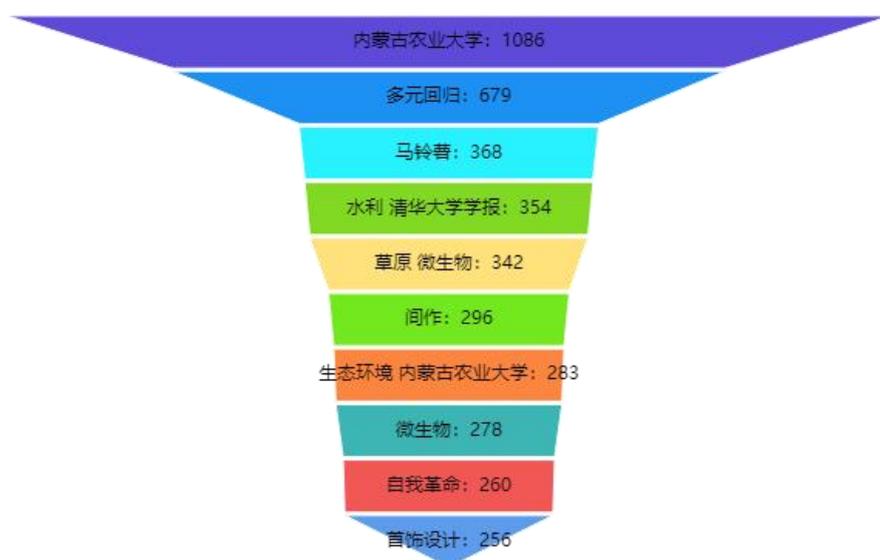


图 2 高频检索词（按照检索次数排序）

从高频检索词分析，可以看到近期学校研究学科领域，包括水利、马铃薯、草原微生物、间作、生态环境等都是学校的重点学科。其中还包括不直接反映学科领域的关键词如清华大学学报、内蒙古农业大学等，通过这些关键词分析也可以间接围绕该用户研究学科领域进行分析。

对于具体高质量、优质文献资源的中国知网文献检索库的资源，可以实行文献下载保留机制，优化馆藏结构。也可以针对特定读者人群推送个性化资源，实现助力我校智慧图书馆的建设。

2.3 检索标题

表 2 标题浏览统计表（TOP20）

标题名称	次数
基于人工智能的计算机大数据安全技术平台构建	214
自然发酵乳制品中细菌多样性的研究进展	159
坚定不移走好高质量发展之路	136
公务员工资该不该涨？	115

荒漠草原区景观动态及生态格局优化研究以内蒙古四子王旗为例	87
新能源汽车技术	71
习近平关于党的自我革命重要论述研究	71
荒漠草原区景观动态及生态格局优化研究 以内蒙古四子王旗为例	67
年产 3000 吨白莲山药凝固型酸奶的工厂设计	66
双碳目标下电力企业财务战略实施效果分析以 H 公司为例	58
水力侵蚀下砒砂岩坡面植被格局对土壤颗粒空间分布的影响	57
筑牢我国北方重要生态安全屏障面临的问题及系统治理方略	56
科学解读反式脂肪酸的危害	52
毕业论文开题报告主体内容的逻辑化撰写	50
国货品牌联名对西安市大学生体育用品消费者购买意愿影响研究以李宁品牌为例	50
鼠李糖乳杆菌 zrx01 冻干菌粉的制备及工艺优化	49
水力侵蚀下砒砂岩区植被格局对微地形和侵蚀的影响	49
一例猫肠梗阻的诊断和治疗	49
外来入侵杂草刺萼龙葵种子萌发机制	48
品牌联名产品属性对消费者购买意愿的影响研究——基于消费者感知价值视角	47

用户在电子资源库中的高频浏览内容，体现了本校读者都在看什么，以全部资源组浏览标题为样本，浏览次数较多的标题有基于人工智能的计算机大数据安全技术平台构建、自然发酵乳制品中细菌多样性的研究进展、坚定不移走好高质量发展之路等。

2.4 热门学科

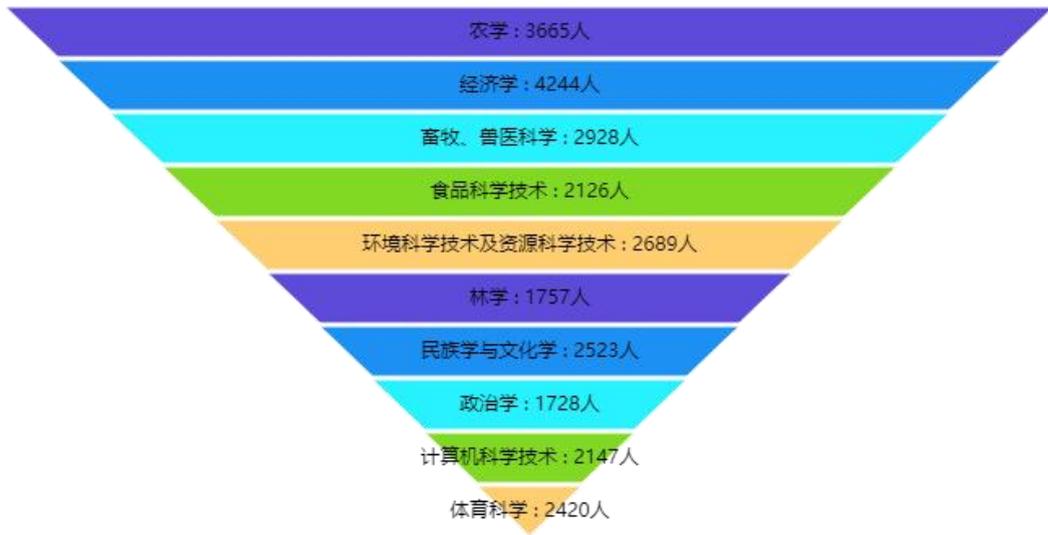


图 3 用户访问一级学科专业表 (TOP10)

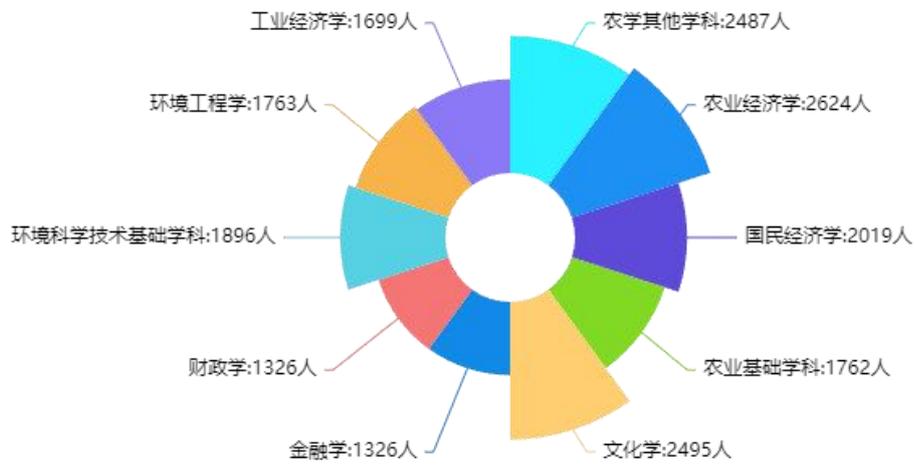


图 4 用户访问二级学科专业表 (TOP10)

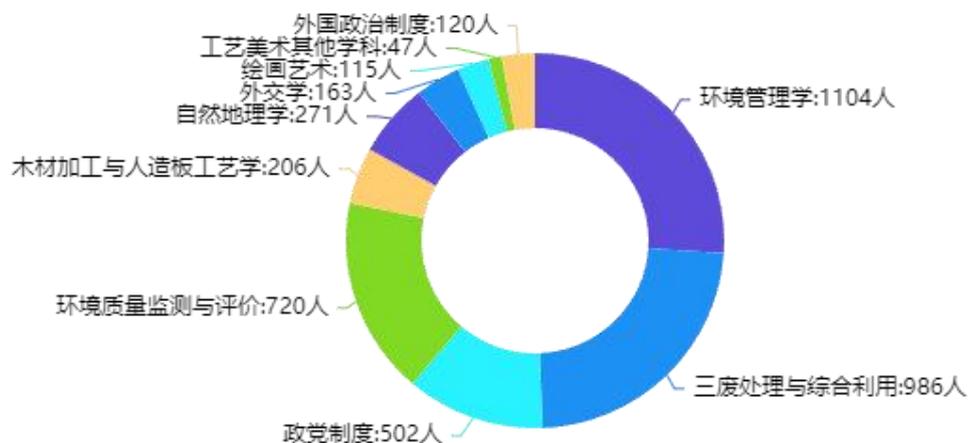


图5 用户访问三级学科专业表 (TOP10)

由读者访问的一级、二级、三级学科分析结果可知，本校师生在农学学科、经济学学科、畜牧兽医学科方面使用人数较多，这些大部分也是学校的重点和双一流学科，图书馆可以根据这些数据及时完善该学科对应的资源配置，带动相关学科和交叉学科的文献资源建设，全面保障学校“双一流”学科和学科群的文献需求。在保障重点学科资源配置的基础上，发掘本校优势学科，查漏补缺，及时申请相关文献资源建设经费，为本校学科教育提供较为持续完善的文献资源保障。

3 行为数据分析

3.1 读者浏览量排名

表3 行为人浏览量排名 (top20)

序号	读者	学籍	读者类型	学院	浏览次数
1	李*磊	2022	硕士生	农学院	4899
2	郝*伟	2022	硕士生	沙漠治理学院	4301
3	曾*择	2022	硕士生	人文社会科学学院	4005

4	高*罕	2022	硕士生	人文社会科学学院	3387
5	王*	2022	硕士生	机电工程学院	3264
6	张*洁	2022	硕士生	经济管理学院	3125
7	徐*悦	2023	硕士生	经济管理学院	3014
8	王*		教职工	动物科学学院	2843
9	苏*潭	2022	硕士生	动物科学学院	2764
10	裴*永		教职工	能源与交通工程学院	2637
11	白*梅	2018	博士生	草原与资源环境学院	2606
12	安*琳	2021	博士生	经济管理学院	2440
13	蔡*	2022	硕士生	农学院	2376
14	桑*华	2022	硕士生	经济管理学院	2328
15	王*博	2023	硕士生	沙漠治理学院	2296
16	李*	2020	硕士生	经济管理学院	2253
17	乌**娃		教职工	人文社会科学学院	2167
18	马*龙	2021	硕士生	马克思主义学院	2108
19	薛*	2022	博士生	经济管理学院	2095
20	李*	2020	本科生	外国语学院	2045

数据库的访问行为人数，直观表达了我校电子图书馆电子资源推广取得的成效，访问次数最多的读者是 2022 级农学院的硕士研究生，行为总量为 4899 次，对现有资源及时进行更新、充实和完善时，就可以有针对性的调研请教该读者的意见，以期吸引更多读者利用数据库，发挥馆藏电子资源的最大效用。在网站宣传馆藏电子资源库时，也可以请特定的读者群体来参与到活动中，来提高活动的实施效果，为数据库的推广提供建设性的建议。

3.2 学院浏览量排名

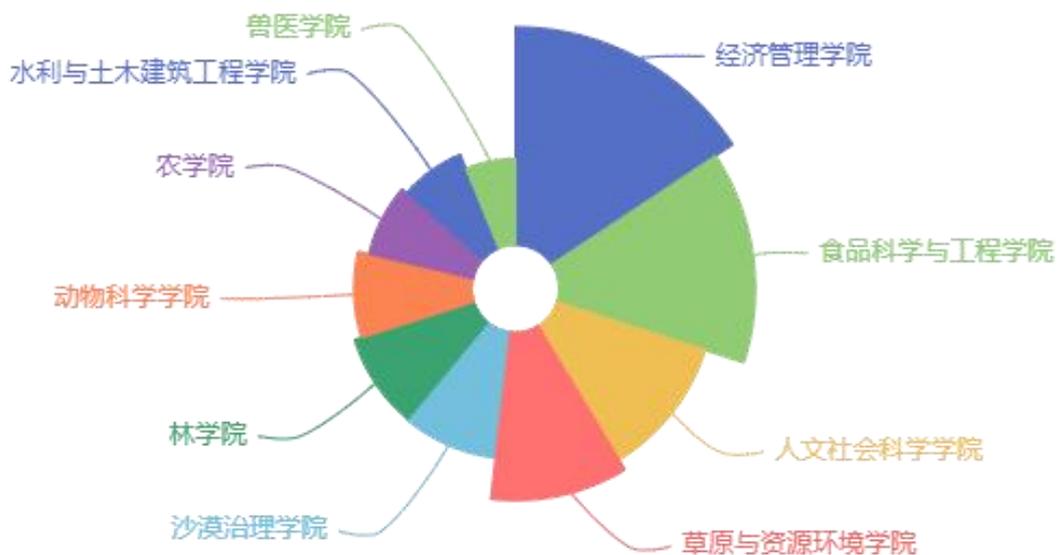


图 6 学院浏览量

表 4 学院浏览量排名 (top10)

学院	行为总量 (次)
经济管理学院	116191
食品科学与工程学院	105409
人文社会科学学院	83674
草原与资源环境学院	78159
沙漠治理学院	67678
林学院	67156
动物科学学院	63422
农学院	57361
水利与土木建筑工程学院	54221
兽医学院	46154

数据库资源利用较多的学院为经济管理学院、食品科学与工程学院、人文社会科学学院、行为总量都 8 万次以上，以后的资源建设方面就可以重点针对这些学院调研，增加相关学科资源总量。

3.3 学级浏览量排名

表 5 学籍访问量

学籍	行为总量 (次)
2022	246026
2020	244251
2021	163347
2023	160743

通过学籍读者使用资源情况可以分析,2022 级读者是最主要的资源使用群体,可以进一步加强对该类读者的信息资源检索培训和宣传。

3.4 读者在线人数趋势

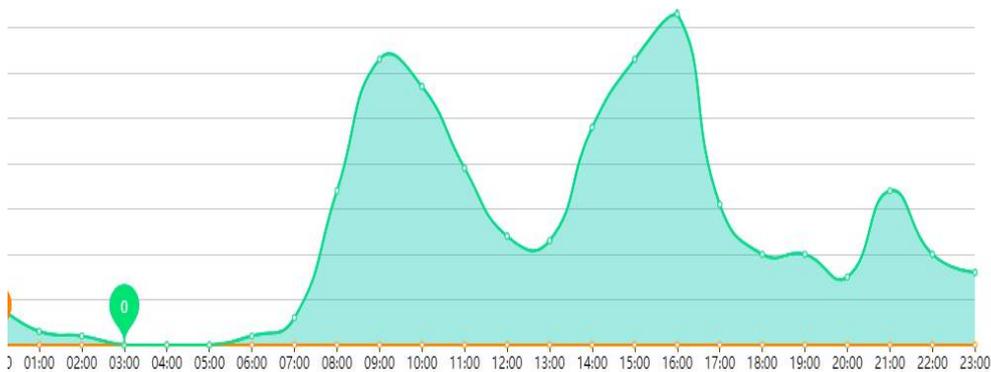


图 7 在线读者量

通过对每日时间段读者访问人数的统计,读者访问高峰期大约是每日上午 9:00-10:00,下午 3:00-4:00。

3.5 读者类型浏览量排名

读者类型访问比例

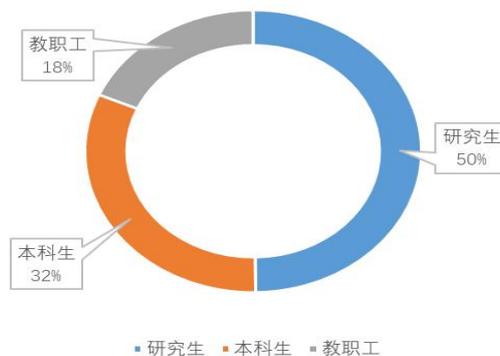


图 8 读者类型访问中国知网比例

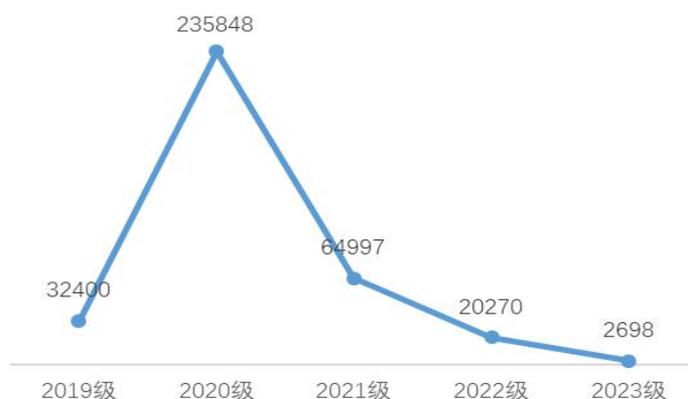


图 9 中国知网数据库访问量（本科生）

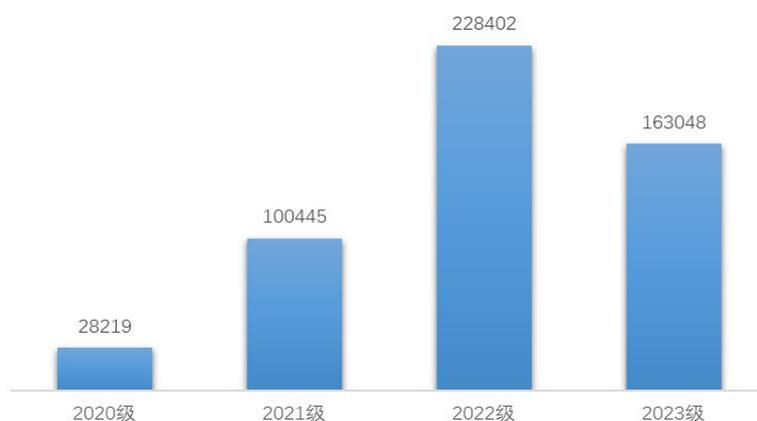


图 10 中国知网数据库访问量（研究生）

根据读者类型读者访问次数可以看出本科生访问人次最多的是 2020 级，研究生访问人次最多的是 2022 级。由以上数据可以看出数据库利用率最高的是研究生。

4 总结及建议

电子资源分析系统能够直观的反应电子资源访问情况，可以为资源试用、采购、使用情况提供一定的数据支撑。

我校共有读者 3 万多人，访问过数据库的人数为 1 万多，有将近三分之二的读者未曾访问过数据库，图书馆可以调查研究是读者没有访问

资源的需求还是资源内容无法满足读者。过后可根据情况针对性的调整资源采购策略，提高电子资源利用率，更好的服务学校学科发展。