



CADAL用户行为挖掘与分析

CADAL项目管理中心 金佳丽

2022.07.22

CONTENTS

1 CADAL项目背景

2 用户行为数据集

3 使用建议

PART ONE

CADAL项目背景

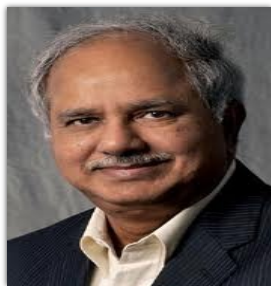


缘起

2000年，中美计算机科学家与图书馆联合发起“中美百万册数字图书馆国际合作计划”（China-America Digital Academic Library，简称CADAL），开始了全球最早的大规模数字化资源工程（Mass Digitization）。



潘云鹤
中国工程院院士



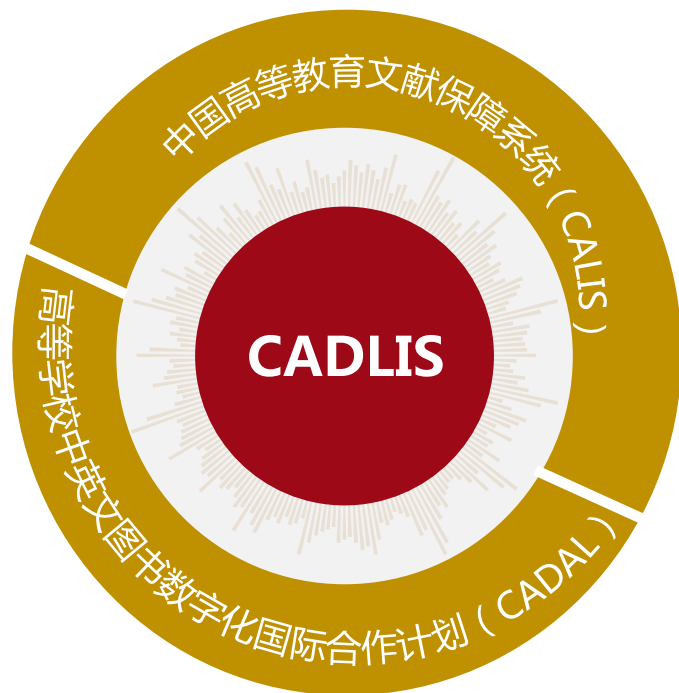
Raj Reddy
计算机图灵奖得主



CADAL进入211工程 公共服务体系

计社会[2002]1505号

2002年9月，国家发改委、教育部、财政部下发《关于“十五211工程”建设的若干意见》，支持“高等学校中英文图书数字化国际合作计划（CADAL）”的实施，为高等学校教学科研提供强有力的数字资源支持。



2009年1月，教育部批复浙江大学立项建设“数字图书馆教育部工程研究中心”。8月，基于“中英文图书数字化国际合作计划（CADAL）”的建设成果，教育部高教司通过了《大学数字图书馆国际合作计划（China Academic Digital Associative Library, CADAL）建设可行性研究报告》CADAL项目二期建设正式启动。

2011年2月25日，“数字图书馆教育部工程研究中心”通过教育部组织的验收。

中华人民共和国教育部

教技函〔2009〕

教育部关于数字图书馆教育部工程研究中心立项的批复



CADAL

数字图书馆教育部工程研究中心
Engineering Research Center of Digital Library, Ministry of Education

浙江大學
2009年1月10日《浙江大學關於申請成立“數字圖書館

教育部工程研究中心立項建設。現批復如

一、該圖書館教育部工程研究中心的技術方向主要是

文化和數字內容服務產業，研發分布式海量跨媒體信息存儲、

和知識挖掘、展示等核心技術，建立數字圖書館相關技術試驗

二、該工程中心的建設目標與任務是：進一步明確工程中

位，以文化和數字內容服務產業相關的工程技術創新和系統

為主要任務，通過向行業內企業的技术轉移與擴散，推動與促

業技術進步和產業發展。積極探索工程中心的運行發展機

发展历程

2000

中美科学家共同倡议，数字化人类1%的知识，即100万册图书。如持之以恒，则10年可以完成。

敢为天下先
We did it first

2001-2006

项目一期
100万册图书
数字化

2007-2012

项目二期，新增150万册图书数字化，标准规范、服务网络、技术服务升级

一览众山小
We do it better

2013-2018

项目运维阶段

- 整合海量资源
- 融合先进技术
- 泛在个性服务
- 全球开放合作

同时开展：数据合作，联盟建设，特藏建设，数字人文

更上一层楼
We can do better

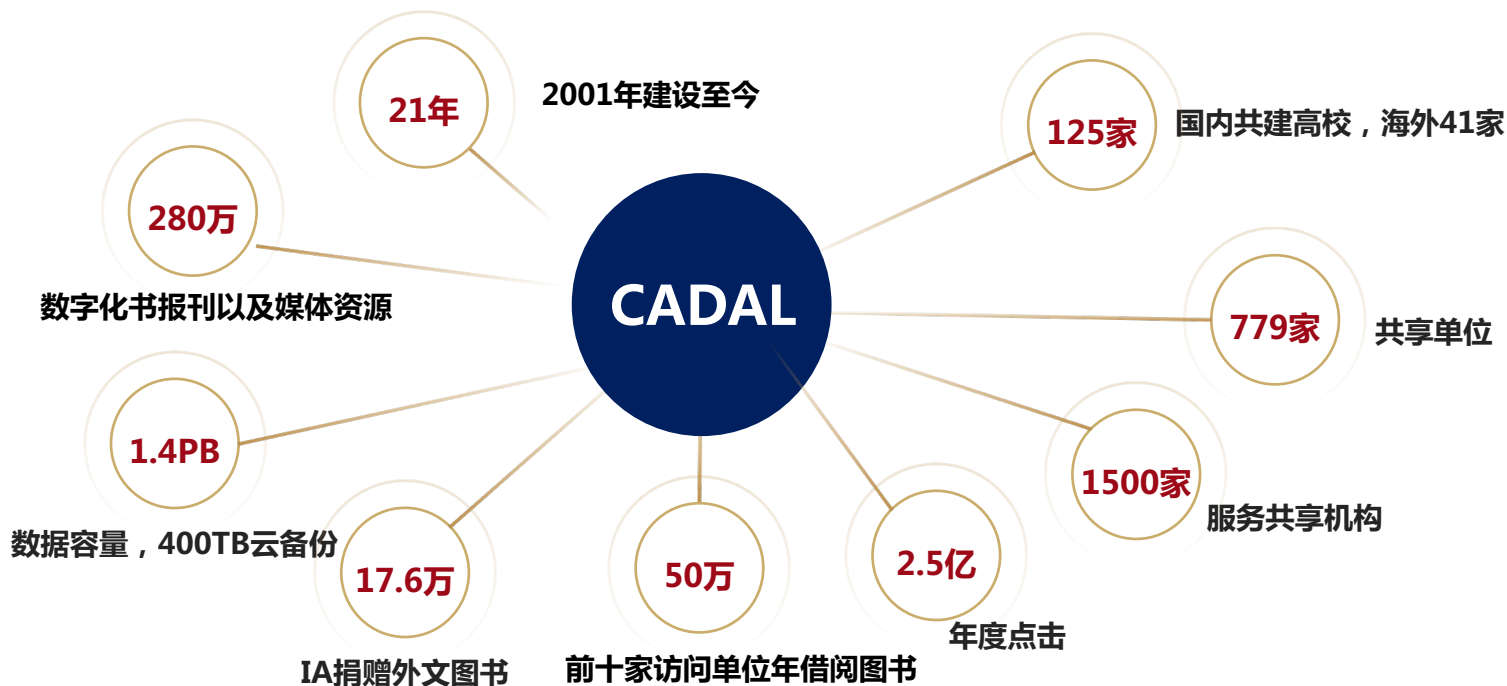
2019-

数据助力知识服务

大数据——大知识

发展现状

CADAL项目的建设把中国高校图书馆由购买资源共享有效推动到了资源的共建共享，建成了拥有超过280万册中英文电子图书的**全球最大公益性数字图书馆**，成为我国**高校图书馆最大规模的知识整合平台之一**。





民国文献大全



当代生活资料



地方志



老照片

特藏库



甲骨数字化



满铁资料



蒋介石资料



墓志拓片



知领特色产品



侨批



哥伦比亚大学图书馆民国文献缩微胶片

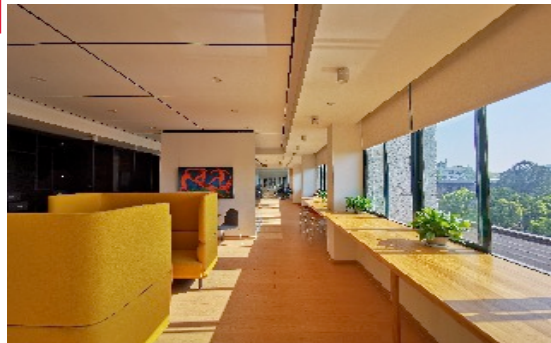


愿景

项目以“共建共享”理念为指导思想，以先进的技术手段，全面整合国内高校图书馆、图书情报服务机构、学术研究机构所拥有或生产的各类信息资源及其相关服务，有重点地引进、共享国际相关机构的各类信息资源与服务。对包括书画、建筑工程、篆刻、戏剧、工艺品等在内的多种类型媒体资源进行数字化整合，项目建成的资源覆盖理、工、农、医、人文、社科等多种学科，通过因特网提供一站式的个性化知识服务，向参与建设的高等院校、学术机构提供教学科研支撑。

使命

构建拥有多学科、多类型、多语种海量数字资源的，由国内外图书馆、学术组织、学科专业人员广泛参与建设与服务，具有高技术水平的学术数字图书馆，成为国家创新体系信息基础设施之一。目前，CADAL已进入项目运维期，现阶段主要致力于保证已建成的项目成果持续有效地发挥作用，并不断增强高等教育公共服务基础设施建设。





用户行为数据集

PART TWO



大学数字图书馆国际合作计划

CHINA ACADEMIC DIGITAL ASSOCIATIVE LIBRARY

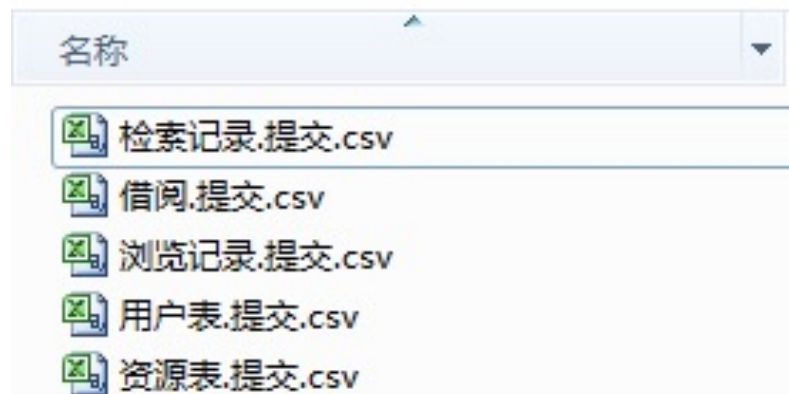
全部 ▾

请输入搜索内容



CADAL

网址：<https://cadal.edu.cn>



| id | user_id | user_name | keyWord | create_time | ip | search_record (检索记录) |
|--------|-----------|-----------|-----------------|----------------|---------------|----------------------|
| 336586 | 52dae6aa6 | 已屏蔽 | 邓小平与改革开放20年重大决策 | 2021/1/1 9:32 | 211.71.28.* | |
| 336587 | 52dae6aa6 | 已屏蔽 | 转型与断裂 改革以来中国社会 | 2021/1/1 9:33 | 211.71.28.* | |
| 336588 | cded8e7c3 | 已屏蔽 | 内蒙古草地资源统计资料 | 2021/1/1 9:46 | 202.205.93.* | |
| 336589 | 6212233a1 | 已屏蔽 | 语法史 | 2021/1/1 9:53 | 123.114.243.* | |
| 336590 | 6212233a1 | 已屏蔽 | 汉语语法史 | 2021/1/1 9:54 | 123.114.243.* | |
| 336591 | 31cb91aa2 | 已屏蔽 | 宋史 | 2021/1/1 9:58 | 42.236.204.* | |
| 336592 | ee9e0c140 | 已屏蔽 | 机械工程测试技术 | 2021/1/1 10:00 | 202.200.234.* | |
| 336593 | 97f058bff | 已屏蔽 | 中学数学课例分析 | 2021/1/1 10:01 | 58.192.87.* | |
| 336594 | a54c7ac37 | 已屏蔽 | 青岛市政府 | 2021/1/1 10:01 | 211.64.159.* | |
| 336595 | fab8fbd85 | 已屏蔽 | 叶适与永嘉学派论文集 | 2021/1/1 10:05 | 219.217.38.* | |
| 336596 | 540689e4c | 已屏蔽 | 课程 民族 | 2021/1/1 10:08 | 202.195.115.* | |

民国经济史

荐购

作者：朱斯煌

出版社：银行学会

馆藏单位：四川大

出版时间：1948-0

资源类型：民国图

标签：民国;经济史

主题：民国;经济

说明：冯子明 汤心

分享 付费

试读

提醒

借阅后有7天阅读时长,本章节剩余 10次

借阅次数,是否借阅此书第 1 章?

暂不借阅

借阅此书

名称

检索记录.提交.csv

借阅.提交.csv

浏览记录.提交.csv

用户表.提交.csv

资源表.提交.csv

| id | userid | username | resourceid | ssno | resourcename | chapter | resource_image | cover_image | author | remain | borrow_start_time | borrow_end_time | ip |
|--------|-----------|----------|------------|----------|---------------------|---------|-------------------------------|-------------|----------------|--------|-------------------|-----------------|---------------|
| 101672 | d43afabef | 已屏蔽 | 33102886 | 33102886 | 中华文化通志·制度文化典(4-037) | | 33/102/886/otiff/00000001. | | 宁欣撰 | | 2021/3/19 9:47 | 2021/3/26 9:47 | 10.15.80.* |
| 101958 | 10862a71f | 已屏蔽 | 6753420 | 6753420 | 雲南省昭通縣志:全 | | 06/753/420/otiff/00000001.jpg | | | | 2021/1/31 20:16 | 2021/2/7 20:16 | 10.15.80.* |
| 102389 | c7a128ada | 已屏蔽 | 33165849 | 33165849 | 真货·假货 | | 33/165/849/otiff/00000001. | | 吕一林 芮利杰 张玉霞 | | 2021/3/25 16:05 | 2021/4/1 16:05 | 101.80.103.* |
| 102640 | 54dc9ac4f | 已屏蔽 | 13292650 | 13292650 | 大辞海·政治学·社会学卷 | | 13/292/650/otiff/00000001. | | 夏征农;陈至立 | | 2021/1/1 0:28 | 2021/1/8 0:28 | 202.119.43.* |
| 102641 | 92c007c1f | 已屏蔽 | 6344328 | 6344328 | 中国文学家大辞典 | | 06/344/328/otiff/00000001. | | 谭正璧(编) | | 2021/1/1 1:24 | 2021/1/8 1:24 | 58.194.168.* |
| 102642 | 92c007c1f | 已屏蔽 | 6514840 | 6514840 | 中国文学家大辞典·清代卷 | | 06/514/840/otiff/00000001. | | 钱仲联(主编) | | 2021/1/1 1:37 | 2021/1/8 1:37 | 58.194.168.* |
| 102643 | 19db35db4 | 已屏蔽 | 33119150 | 33119150 | 唐五代画论 | | 33/119/150/otiff/00000001. | | 何志明 潘运告编著 | | 2021/1/1 2:10 | 2021/1/8 2:10 | 59.66.104.* |
| 102644 | fa34f302f | 已屏蔽 | 7109283 | 7109283 | 振动结构模态分析 | | 07/109/283/otiff/00000001. | | 曹树谦(编著);张文德(译) | | 2021/1/1 2:21 | 2021/1/8 2:21 | 219.217.246.* |
| 102645 | cd8ea36d2 | 已屏蔽 | 16000632 | 16000632 | 金陵大学六十周年纪念册 | | 16/000/632/otiff/00000001. | | 金陵大学 | | 2021/1/1 4:12 | 2021/1/8 4:12 | 116.1.242.* |

名称

检索记录.提交.csv

借阅.提交.csv

浏览记录.提交.csv

用户表.提交.csv

资源表.提交.csv

银行界二十周年纪念刊
 民國經濟史
 林冲 冯子明 汤心儒 鲍文熙

民国经济史

荐购

作者：朱斯煌

出版社：银行学会银行通报社·上海

馆藏单位：四川大学

出版时间：1948-01(民国三十七年)

资源类型：民国图书

标签：民国;经济史;上海;四十年代;民国三十七年;专著 [添加标签](#)

主题：民国；经济史；金融

说明：冯子明 汤心儒 鲍文熙

[分享](#) [付费打印](#) [我是作者](#) [捐赠](#) [完善](#)

试读

章节阅读

收藏

| id | userid | username | resourceid | resourcechapter | resource | i:cover | img | author | browsing_datetime | type | ip |
|---------|-----------|----------|------------|-----------------|---------------------|---------|-----|--------|-------------------|------|---------------|
| 2352349 | 54dc9ac45 | 已屏蔽 | 13292650 | 大辞海·政治学·社 | 13/292/650/otiff/00 | | | 夏征农; 陈 | 2021/1/1 0:23 | | 202.119.43.* |
| 2352350 | 54dc9ac45 | 已屏蔽 | 13292650 | 大辞海·政治学·社 | 13/292/650/otiff/00 | | | 夏征农; 陈 | 2021/1/1 0:27 | | 202.119.43.* |
| 2352351 | 8b62e0229 | 已屏蔽 | 59008220 | 数值方法和MATLAB实 | 59/008/220/otiff/00 | | | 拉克唐瓦尔 | 2021/1/1 0:44 | | 202.200.234.* |
| 2352352 | 54dc9ac45 | 已屏蔽 | 13292650 | 大辞海·政治学·社 | 13/292/650/otiff/00 | | | 夏征农; 陈 | 2021/1/1 1:01 | | 202.119.43.* |
| 2352353 | 92c007c19 | 已屏蔽 | 6344328 | 中国文学家大辞典 | 06/344/328/otiff/00 | | | 谭正璧(编 | 2021/1/1 1:22 | | 58.194.168.* |
| 2352354 | 92c007c19 | 已屏蔽 | 6514840 | 中国文学家大辞典· | 06/514/840/otiff/00 | | | 钱仲联(主 | 2021/1/1 1:24 | | 58.194.168.* |
| 2352355 | 92c007c19 | 已屏蔽 | 6514840 | 中国文学家大辞典· | 06/514/840/otiff/00 | | | 钱仲联(主 | 2021/1/1 1:26 | | 58.194.168.* |
| 2352356 | fa34f3027 | 已屏蔽 | 7109283 | 振动结构模态分析 | 07/109/283/otiff/00 | | | 曹树谦(编 | 2021/1/1 2:21 | | 219.217.246.* |

名称

检索记录.提交.csv

借阅.提交.csv

浏览记录.提交.csv

用户表.提交.csv

资源表.提交.csv

| id | username | password | status | role | ireal | ibalar | head | rbook | com | book | hic | affiliation | email/邮箱 | sex/性别 | birth | hometown/ | ihab | school/学校 | major/ | ite | bri | registration_date/ | acc | paint | backgrou | recently_login/最近登 | seeti |
|------|----------|----------|--------|------|-------|--------|------|-------|-----|------|-----|-------------|------------|--------|-------|-----------|--------|-----------|--------|--------|--------------------|--------------------|-----|-------|------------------|--------------------|-------|
| 22e | 已屏蔽 | 已屏蔽 | 1 | 已屏蔽 | 0 | 已屏蔽 | yes | yes | yes | yes | yes | *****@163. | female | ### | 辽宁-沈阳 | | 北京师范大学 | 学生 | | | 2011/10/24 23:12 | | 0 | 已屏蔽 | 2021/4/12 9:15 | | |
| 146f | 已屏蔽 | 已屏蔽 | 1 | 19 | 已屏蔽 | 0 | 已屏蔽 | yes | yes | yes | yes | 内蒙古农业大 | *****@im | male | ### | 山东-济南 | | 山东大学 | 学生 | 185*** | 2011/10/31 18:17 D | 0 | 0 | 已屏蔽 | 2021/8/29 16:20 | | |
| b81a | 已屏蔽 | 已屏蔽 | 1 | 已屏蔽 | 0 | 已屏蔽 | yes | yes | yes | yes | yes | 暨南大学 | *****@jnu. | male | ### | 其他-其他 | | 暨南大学 | 学生 | 159*** | 2011/11/17 8:54 | 0 | 0 | 已屏蔽 | 2021/3/25 16:51 | | |
| 284f | 已屏蔽 | 已屏蔽 | 1 | 已屏蔽 | 0 | 已屏蔽 | yes | yes | yes | yes | yes | 四川大学 | *****@126. | female | ### | 其他-其他 | | 北京师范大学 | 学生 | 159*** | 2011/11/18 13:47 | 0 | 0 | 已屏蔽 | 2021/11/24 11:14 | | |
| 404f | 已屏蔽 | 已屏蔽 | 1 | 已屏蔽 | 0 | 已屏蔽 | yes | yes | yes | yes | yes | 内蒙古大学 | *****@qq. | female | ### | 广东-广州 | | 暨南大学 | 历史学 | 156*** | 2014/1/10 13:45 | 0 | 0 | 已屏蔽 | 2021/8/20 9:39 | | |
| 31e | 已屏蔽 | 已屏蔽 | 1 | 18 | 已屏蔽 | 0 | 已屏蔽 | yes | yes | yes | yes | 温州大学 | *****@qq. | male | ### | 湖北-武汉 | | 温州大学 | 教师 | 159*** | 2011/12/6 18:57 D | 0 | 0 | 已屏蔽 | 2021/11/8 17:39 | | |

名称

- 检索记录.提交.csv
- 借阅.提交.csv
- 浏览记录.提交.csv
- 用户表.提交.csv
- 资源表.提交.csv

| sopath/图书封面地址 | pagenu | cno | title/题名 | title_old/原题名 | creator/作者 | creator_old/原作者 | subjsubje | description/摘要 | description_old/摘要 |
|----------------------|--------|--------|------------------|------------------|---------------------|---------------------|-----------|------------------|--------------------|
| 761/002/777/otiff/00 | 190 | 500111 | 御纂朱子全書：卷四十九至卷五十 | 御纂朱子全書：卷四十九至卷五十 | 宋朱熹（著） | （宋）朱熹（著） | | 本書包括卷一~卷六十六。 | 本書包括卷一~卷六十六。 |
| 761/002/541/otiff/00 | 215 | 200635 | 四書大成：第四冊 | 四書大成：第四冊 | 清沈磊等（纂訂）；清陸摺等 | （清）沈磊等（纂訂）；（清）陸摺等 | | 本書包括第一冊~第五冊。原書部分 | 本書包括第一冊~第五冊。 |
| 761/002/634/otiff/00 | 86 | 701212 | 宋學士全集：卷二十三 | 宋學士全集：卷二十三 | （明）宋濂（著） | （明）宋濂（著） | | 本書包括卷一~卷三十二。 | 本書包括卷一~卷三十二。 |
| 761/002/075/otiff/00 | 103 | 700744 | 百美新詠：第三冊 | 百美新詠：第三冊 | | 不詳 | | 本書包括第一冊~第三冊。 | 本書包括第一冊~第三冊。 |
| 761/002/573/otiff/00 | 106 | 700032 | 宋詩鈔選四集·宋詩鈔三集：第二冊 | 宋詩鈔選四集·宋詩鈔三集：第二冊 | （清）吳孟舉（輯）；（清）吳自牧（輯） | （清）吳孟舉（輯）；（清）吳自牧（輯） | | 本書包括第一冊~第十冊。原書中縫 | 本書包括第一冊~第十冊。 |
| 761/002/640/otiff/00 | 92 | 701212 | 宋學士全集：卷二十九 | 宋學士全集：卷二十九 | 明宋濂（著） | （明）宋濂（著） | | 本書包括卷一~卷三十二。原書部分 | 本書包括卷一~卷三十二。 |
| 761/002/972/otiff/00 | 95 | 703812 | 皕鶴堂批評第一奇書金瓶梅：卷十六 | 皕鶴堂批評第一奇書金瓶梅：卷十六 | （明）蘭陵笑笑生（撰）；（清）張竹坡 | （明）蘭陵笑笑生（撰）；（清）張竹坡 | | 本書包括卷首~卷二十。原書中縫緊 | 本書包括卷首~卷二十。 |
| 761/002/880/otiff/00 | 133 | 200020 | 五經大全·禮記：卷五 | 五經大全·禮記：卷五 | （清）來爾繩等（纂輯） | （清）來爾繩等（纂輯） | | 本書包括卷一~卷十。 | 本書包括卷一~卷十。 |



PART THREE

使用建议



本数据集涵盖2021年CADAL网站完整的用户行为数据，涉及用户信息、资源信息、检索信息、借阅信息、浏览信息相关字段150个。对探寻高校读者电子资源阅读动态、基于读者需求的电子资源采购、基于阅读行为的个性化推荐服务、基于读者交互行为的网站功能评估、不同院校/专业读者的电子书阅读喜好研究等众多方向都具有丰富的研究和应用价值。例如，可以使用数据分析软件或自主编程进行如下分析：

- （1）构建用户阅读素养模型。基于用户行为数据进行建模，以更加精准科学的方式对用户阅读进行分析和解读；
- （2）进行个体/群体偏好分析，如浏览、借阅、评论、收藏，根据用户的阅读行为进行个性化推荐服务；
- （3）进行用户共建参与度分析，如：捐赠副本、完善资源信息；



THANK YOU

谢谢！