

激活数据要素价值： 构建高校图书馆新型数据治理体系

天津 2026.4.2 魏铭

2026年中国图书馆数据治理与应用创新研讨会

第一章

数据价值演变

2026年中国高校图书馆数据治理与应用创新研讨会



1.数据语料成为个人学术研究刚需

基于对100篇2024-2026年文献的系统分析[C1-C85, E1-E25], 高校图书馆服务教学科研, 需要做好以下核心准备和调整:

9.1 战略层面的调整

- 制定AI发展战略: 明确目标、路径和资源配置[C38, E18]
- 建立跨部门协作机制: 教学、科研、技术部门联动[C92]
- 构建开放创新生态: 产学研用协同创新[C94]

9.2 组织层面的准备

- 组建专业团队: 技术+服务+伦理复合型人才[C35]
- 全员能力提升: 系统化AI素养培训体系[C26, C33]
- 调整服务流程: AI融入核心业务流程[C31]

9.3 技术层面的建设

- 夯实数据基础: 高质量数据治理体系[C24]
- 建设服务平台: 可扩展的AI服务架构[C31]
- 确保安全伦理: 负责任的AI应用框架[C31, E1]

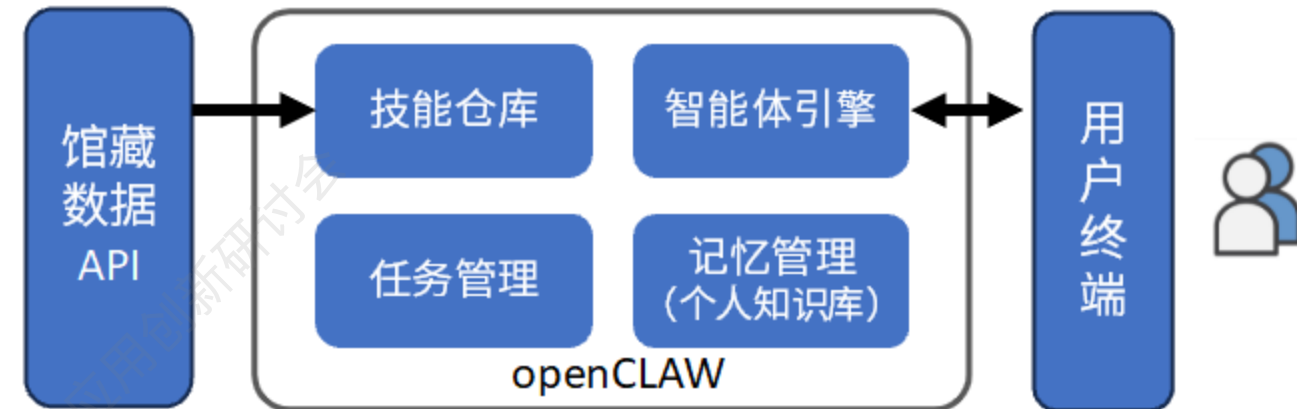
9.4 服务层面的优化

- 教学支持深化: 个性化、智能化学习支持[C41, C49]
- 科研支持升级: 全流程智能科研辅助[C25, C50]
- 用户体验提升: 人性化、无障碍服务设计[C75, C82]

9.5 治理层面的完善

- 建立伦理委员会: AI应用伦理审查[C31, C96]
- 制定评估体系: 数据驱动的效果评估[C74, C89]
- 构建改进机制: 持续迭代优化的闭环[C55, C100]

openCLAW (龙虾) 接入10亿文献数据语料后, 经过用户几轮训练, 检索结果与深度研究的生成质量已显著优于通用大模型



知识服务范式的跨越性转变

随着AI技术的持续迭代, 智能体从被动响应演进为自主决策, 每个用户都能轻松构建自己的“超级智能体”。将改变用户获取知识的习惯与路径。用户更需要质量的学术数据语料。

图书馆面向读者的知识服务, 必然经历范式迁移:



2.数据语料是加速科学发现的燃料

标志性案例：大圣科研合伙人

上海科学智能研究院与复旦大学联合发布的"大圣"科研合伙人，在心脏病研究中展现出惊人能力：AI自主从5亿篇文献中精准定位目标，并行协作完成11轮代码迭代，性能提升11%，将研究生半个月的工作压缩至分钟级完成。

5亿篇
AI自主检索文献规模

分钟级
替代半月人工工作量

国家"十五五"规划明确提出对数字AI科研能力的战略要求，高质量数据已被列为AI for Science学术范式变革的核心基础设施。



从5亿文献中精准定位高度相关的

未命名

资源

+ 添加数据源

请输入科学问题

科研搜索

所有来源 已选择20个数据源

- Deep learning in single-cell and spa...
- Transfer learning enables prediction...
- scQNN is a novel graph neural netwo...
- scDeepSort: a pre
- Transformer for or
- Deep learning app
- Deep-learning-ba
- Deep learning tact
- Deep learning of s
- Inferring gene reg
- SIGNET: single-ce
- Deep learning-bar

对话

文献负责人

在你关注的“心肌梗病 (Cardiomyopathy) + 深度学习基础模型”交叉方向中，目前明确有应用前景的顶级期刊模型主要是：

- ✅ **Geneformer (Nature, 2023; 引用837次)**

- 直接应用于心脏病：原文明确指出，该模型在微调后成功用于扩张型心肌病 (DCM) 患者有限样本中，识别出关键调控因子 (如 *78X20*、*NKX2-5*) 作为候选治疗靶点。
- 机制：通过预训练获得的基因共表达与调控网络知识，在稀有疾病数据下仍能有效迁移。
- 适用性最强：是你当前任务最直观、最权威的参考。

其他模型虽在心脏病中直接验证，但具备基础模型潜力：

提出优化方案

优化结果

原代码

优化代码

自主完成设计方案 写代码 跑实验 总结报告的全过程

对方案进行评估

修改代码

3.数据语料是智能决策的分析引擎

将大模型接入图书馆运行数据后，智能体能够自动挖掘数据趋势，生成的数据报告可充分呈现读者服务的亮点与改进要点，实现从"人工出报表"到"智能洞察"的跨越。

智能问数核心能力

流程减负 自动化生成周报、月报、年报，释放馆员精力

决策支撑 基于真实行为数据的科学决策，替代经验判断

服务洞察 精准识别读者需求变化，驱动服务创新

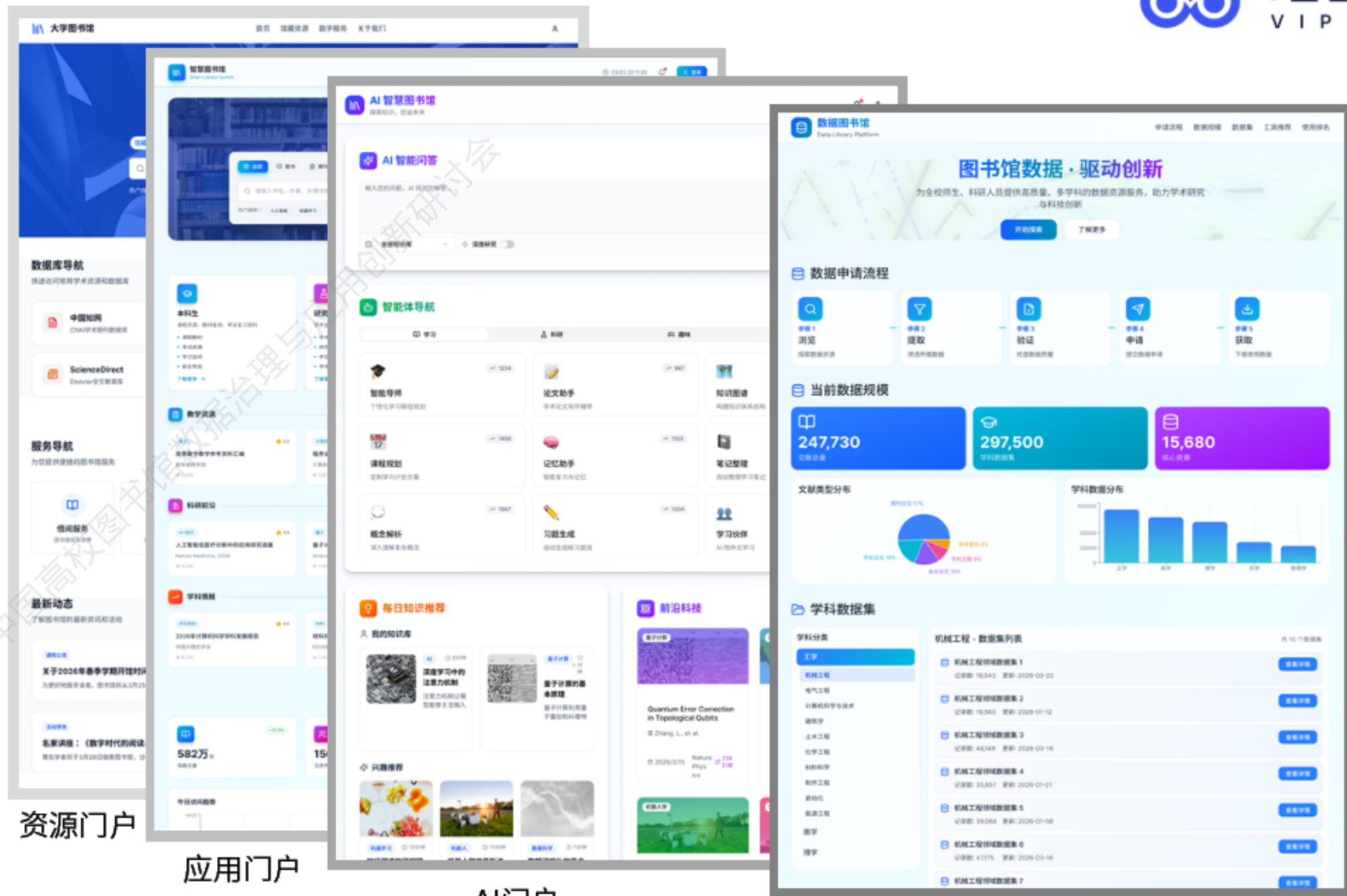


4. 数据即图书馆服务：DaaS

今天的数据价值已不再是藏于服务背后的“生产原料”。

而是能够将图书馆知识充分挖掘、把数据形成安全、便捷、可直接调用的服务产品，是能产生价值的**数据要素**。

提供数据服务的图书馆将更受读者青睐，在高校知识生态中占据核心枢纽地位。



资源门户

应用门户

AI门户

数据门户

2026年中国高校图书馆数据治理与知识创新研讨会

5. 高质量数据基座：是语料（数据要素）的根基

资源 > 数据 > 语料（数据要素）

只有进行科学的数据治理才能支撑图书馆构建新型服务体系的数据基座。

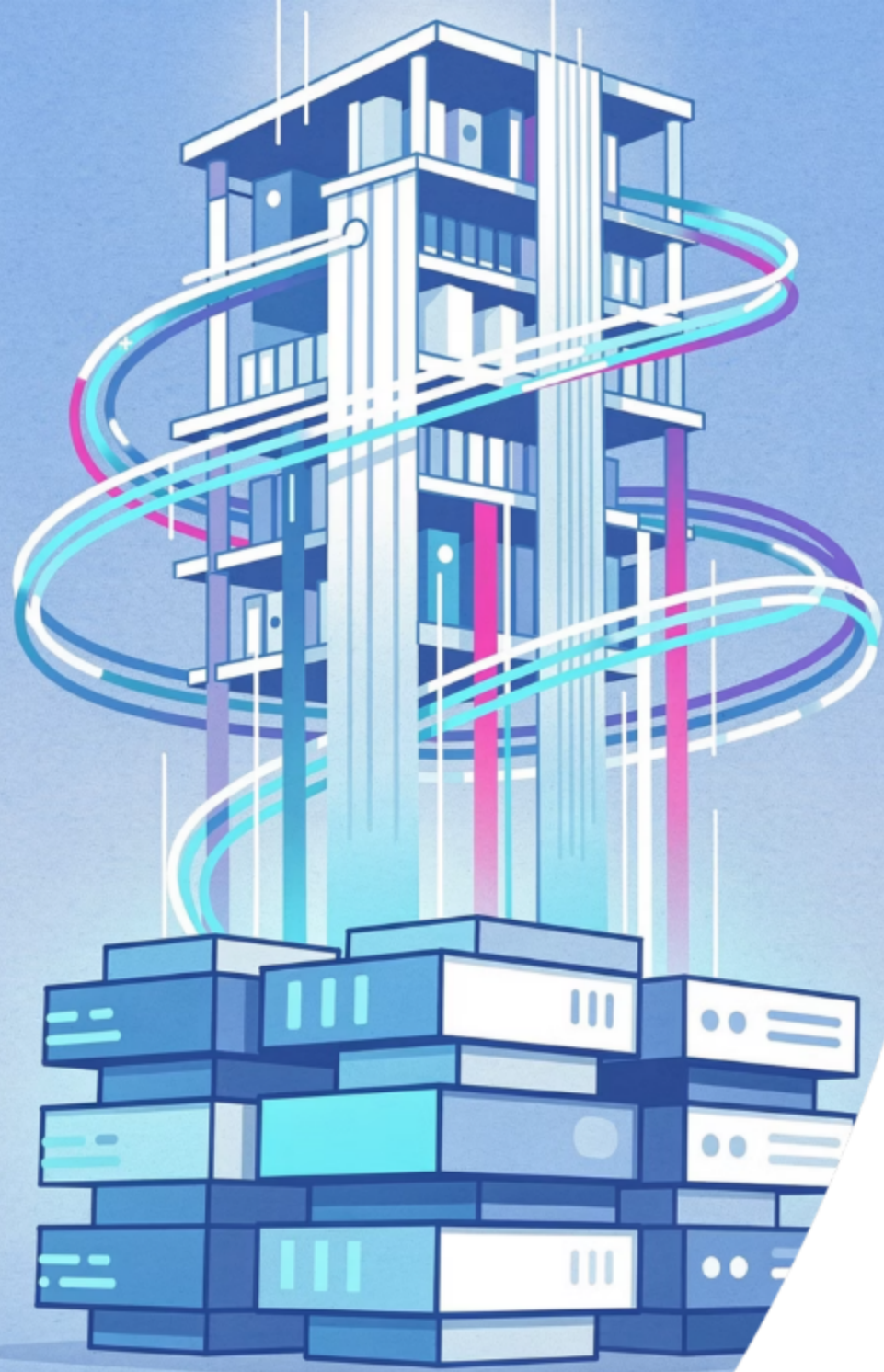
- ❑ 数据治理是将零散数据转化为可信、可用、可持续的战略资产，进而驱动服务创新与价值释放。

竞争格局判断

在AI时代，能更好地治理和利用数据，就能在知识服务领域占据领先地位。

高质量数据基座是未来图书馆服务能力的底层支撑，也是差异化竞争的护城河。

2026年中国高校图书馆数据应用创新研讨会



6. 传统数据体系 VS 新型数据体系

数据治理效果的好坏，不在于“有没有数据”，而在于“能不能服务”。

治理维度	传统数据治理体系（LSP/自研数据中台）	新型数据治理体系
建设目标	管理决策、资源服务支撑	管理决策、知识服务支撑、嵌入教学科研
数据范围	资源数据（图书、期刊）、行为日志、用户数据	多模态元数据（书、刊、论文、专利等）、馆藏状态、全文定位地址、读者服务日志（入馆、借阅、检索、下载...）、业务对象（用户、数据库、设备、事务）、学科与评价标识、引文数据、实体对象（学者、机构、基金..）
数据粒度	篇级 / 册级	篇级 / 册级 / 条目级 / 语料级
实时性	批量处理	核心数据准实时 / 文献数据日级更新
权限管理	按角色/部门授权；敏感信息默认不可见	按角色/部门/场景授权；敏感信息默认不可见；对外共享要可控
隐私与合规	符合法律对个人信息处理的合法、必要要求；满足网络安全与合规审计要求	符合法律对个人信息处理的合法、必要要求；满足网络安全与合规审计要求
可用性	可找、可取；	可找、可取、可用；指标口径一致；
API/接口	API、中间库	API、中间库、MCP、Skill

第二章

治理方法

DALIB数据体系建设方案

2026年中国高校图书馆数据治理与应用创新研讨会



1. DALIB 3.0 定位：图书馆新型数据治理解决方案

数据治理是一项系统性、专业性的工程，其核心目标是确保数据的完整性、准确性、规范性与时效性。

DALIB：采用成熟的数据治理模型，专为图书馆打造的收集、治理、质检、抽取、统计、监控等数据全生命管理平台

3.0 版本升级：从数据治理到数据驱动

数据接入与治理

统计与分析

应用套件

文献中心

指标中心

开放平台

馆藏绩效

运行中心

数据空间

数据大屏

阅读报告

馆藏数据仓库

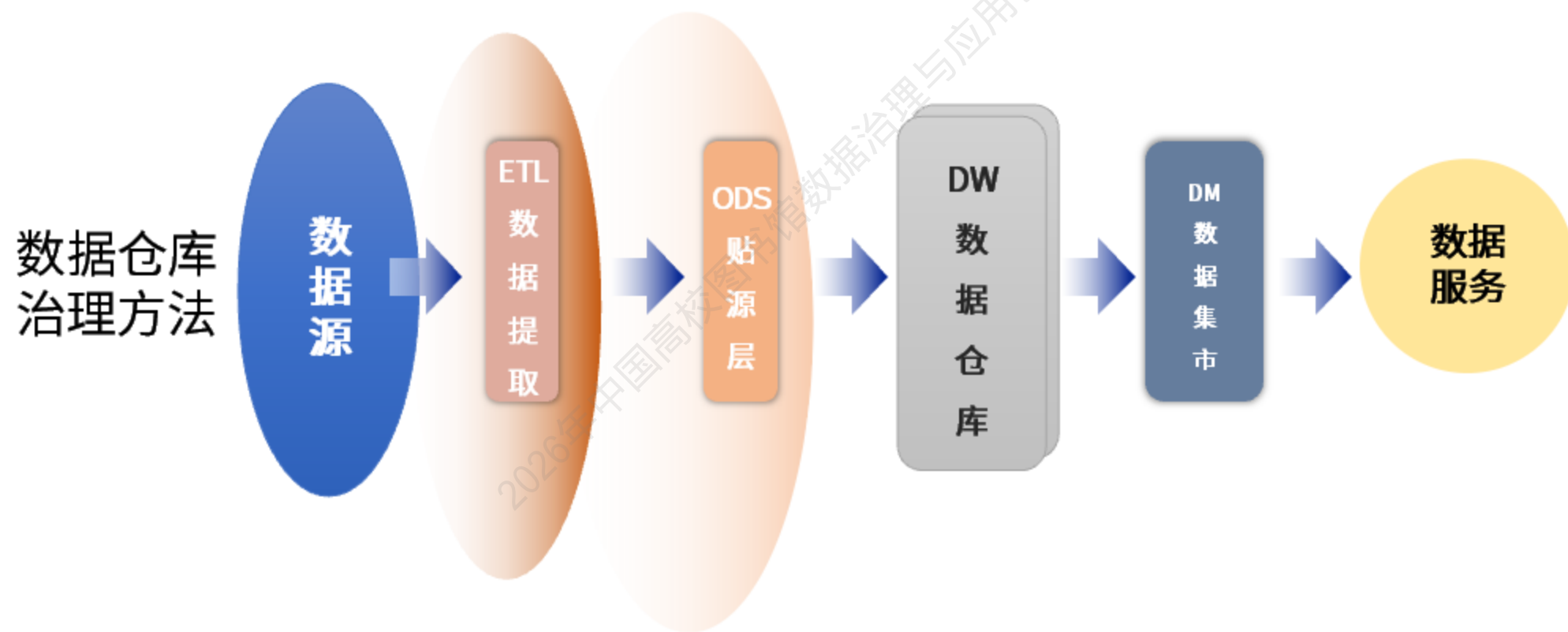
知识库

DALIB3.0模块框架

2. 治理方法：工业化成熟的数据治理模型

图书馆的数据环境极其复杂，涵盖异构文献数据库、行为日志与业务管理数据。

DALIB 3.0 采用全新的技术架构，基于**多层数据治理模型**（ODS → DW → DM → ADS），将数据治理划分为多个层级协同推进。确保数据治理的**完整性、准确性与稳定性**。每层均设有质量阀控与血缘机制，支持可回溯核对。



新的治理框架，将数据准确率提升至**98%**以上，数据完整性提升至**99%**以上，为后续的创新服务提供坚实支撑。

3. 数据治理标准：双向扩展

数据标准决定了数据治理框架的范围、质量要求与数据可服务的场景边界。标准的高度决定服务的深度。

文献数据治理深度延伸

覆盖主流文献类型

图书、期刊、学位论文、会议论文、专利、标准、报告、报纸、视频等全覆盖，支持自定义文献类型扩展

多分类标注体系

中图学科分类与教育部学科分类并行标注，支持自定义分类设定。满足不同场景的分类聚合与分析需求

实体对象关联

作者、机构、基金、期刊等实体关联标注、中外核心引文关系标注。满足研究关系分析需求。

编号	字段	说明
1	DC.Identifier	ID
2	DC.Title	标题
3	DC.Title.Subhead	副标题
4	DC.Title.Alternative	外文标题
5	DC.Creator	作者
6	DC.Creator.Alternative	作者外文名
7	DC.Creator.Correspond	通信作者
8	DC.Creator.Correspond.Address	通讯作者地址
9	DC.Creator.Email	作者邮箱
10	DC.Creator.Information	作者信息
11	DC.Creator.Organ	作者机构
12	DC.Creator.Organ.Alternative	作者机构外文名

40+初始字段
篇级数据粒度

分类名	层级	分类号
哲学	门类	01
哲学	一级	0101
经济学	门类	02
理论经济学	一级	0201
经济思想史	二级	020102
经济史	二级	020103
西方经济学	二级	020104
应用经济学	一级	0202
法学	门类	03
教育学	门类	04
文学	门类	05
历史学	门类	06
理学	门类	07

学科分类、中图分类
中外分类与学科映射

2026年中国高校图书馆数据治理与应用创新研讨会

4.数据治理标准：文献收录指标体系



DALIB 3.0 在文献标准中全面增加学术评价指标的完整标注，让文献数据都具备清晰的"学术权重"，为科研评价与学科服务提供权威数据依据。

重要收录	版本
JCR期刊	JCR2024
ESI来源刊	ESI202509
EI来源刊	EI202510
SCOPUS来源刊	SCOPUS202510
中科院分区表来源刊	CAS2025
北大核心期刊	BDHX2023
CSSCI期刊	CSSCI2023_2024
CSCD期刊	CSCD2023_2024
SCIE来源刊	SCI202510
SSCI来源刊	SSCI202510
A&HCI来源刊	AHCI202510

重要收录指标

学科分类编码	0703;070304;0817;081704
学科分类名称	化学;物理化学;化学工程与技术;应用化学
简介	是基础学科类学术刊物，由中国科学技术协会主管、中国化学会和北京大学共同主办、北京大学化学学院物理化学学报编辑部编辑出版。月刊。主要刊载化学学科物理化学领域具有原创性实验和基础理论研究类文章。
主管部门	中国科学技术协会
cas_small_name	CHEMISTRY, PHYSICAL 物理化学
cas_class_num	2@2025;2@2023;2@2022;3@2019;3@2020;3@2021
jcr_part_name	CHEMISTRY, PHYSICAL:SCIE:PHYSICAL:SCIE:Q1@2022;CHEMISTRY, PHYSICAL:SCIE:Q1@2024;CHEMISTRY, PHYSICAL:SCIE:Q1@2023;CHEMISTRY
jcr_if_rank	CHEMISTRY, PHYSICAL:136/159@2019;CHEMISTRY, PHYSICAL:53/165@2021;CHEMISTRY, PHYSICAL:52/163@2021;CHEMISTRY, PHYSICAL:20/185@2024;CHEMISTRY, PHYSICAL:25/161@2022;CHEMISTRY, PHYSICAL:110/134@2011;CHEMISTRY, PHYSICAL:24/178@2023
jcr_jci	0.45@2021;1.34@2024;N/A@2013;N/A@2014;N/A@2015;N/A@2016;1.12@2023;
jcr_jci_rank	CHEMISTRY, PHYSICAL:153/168@2020;CHEMISTRY, PHYSICAL:43/178@2023;CHEMISTRY, PHYSICAL:31/185@2024;CHEMISTRY, PHYSICAL:157/164@2018;CHEMISTRY, PHYSICAL:146/156@2017;CHEMISTRY, PHYSICAL:109/174@2021;CHEMISTRY, CHEMISTRY, PHYSICAL:51/172@2022;CHEMISTRY, PHYSICAL:107/172@2021
jcr_part_num	Q1;Q1@2024;Q1@2023;Q4@2003;Q1@2022; Q4@2019;Q4@2016;Q4@2017
esi_rank	CHEMISTRY@214
scopus_part_name	Physical and Theoretical Chemistry:Q1@2024;Physical and Theoretical Chemistry:Q1@2023
scopus_part_num	Q1;Q1@2024;Q1@2023
scopus_sjr	1.929@2024;1.606@2023
h_index	47
vip_impact	2.0672@2024;1.9875@2023;1.6797@2022;1.2183@2021;0.5887@2020
vip_impact_latest	2.0672
cnki_zh_impact	3.819@2025;3.389@2024;2.892@2023;1.374@2023;1.374@2022;0.658@2021;0.674@2020;0.609@2019;0.791@2018;1.423@2017

重要评价指标

5. 数据治理标准：加入业务与维度数据实现全域数据融合



DALIB 3.0 将运行数据从单一的“日志统计”全面升级为涵盖行为日志、业务流程数据与公共维度属性的三元融合体系，真正实现全域数据的统一治理与深度利用。

行为数据标准

默认提供 **28种** 行为数据标准类型，覆盖入馆、借阅、检索、下载、预约、续借等全链路读者行为，并支持各馆根据实际需求 **自定义行为数据标准**，确保本地业务特色得到完整记录。

业务数据维度

支持用户、学院专业、数据库、学科等默认维度，构建多维数据分析能力，同时支持自定义维度扩展，满足不同高校的管理分析场景，让数据统计口径真正与机构实际吻合。

行为数据

类型名称	编码	类型名称	编码	类型名称	编码	类型名称	编码
门禁入馆	10001	座位签到	10031	链接跳出	10036	检索下载	10008
门禁出馆	10002	自助打印	10034	功能操作	10037	直接下载	10010
空间预约	10013	在线浏览	10015	数据埋点	10038	文献传递	10027
座位预约	10030	登录认证	10003	文献检索	10019	图书借出	10006
空间签到	10032	认证请求	10004	文献浏览	10021	图书续借	10039

28种常见行为标准，支持自定义扩展

业务数据

- 新闻发布
- 活动参与
- 咨询回复
- 数据库信息
- 读者信息
- 馆员信息
- 馆舍面积
- 研讨室
- 服务器设备
- 图书定位
- 环境温度
-

用户、数据库、活动、咨询等

6. 数据集成：多维度保障数据获取效率

合作生态

支持主流厂商原生集成，覆盖国内外主要图书馆系统供应商。

中央知识库

提供亿级数据储备，为数据治理与知识服务提供坚实的基础数据支撑。

ETL 多类型对接

支持多种数据格式与传输协议，实现跨系统数据的标准化抽取、转换与加载。



7. 数据集成：管理制度是根本

管理命题：建立数据存缴制度与规范

再强的技术手段，也依赖可靠的数据来源。图书馆需要有意识地将数据存缴形成管理办法，纳入采购标准与合同条款：



数据库商

采购合同中明确要求免费提供结构化元数据，作为合作前提条件



系统厂商

所有新购系统须提供标准数据接口，确保业务数据可采集可治理



以重庆大学图书馆为例：每年新购系统均提供数据接口，新购数据库提供元数据。每年新增十多条数据对接——这正是在采购环节对厂商提出的明确要求所带来的成果。



8. 数据处理：技术领先，保障数据质量

多来源去重：专利级关系网络识别

基于自主研发专利技术的关系网络识别算法，相比传统规则合并方式，去重准确性有效提升 **50%**，显著降低数据冗余与质量风险。

多来源关联：利用数据补充数据

通过文献数据合并、行为数据合并与交叉数据合并，以数据修复数据、以数据扩展数据，构建更完整的知识图谱。

大模型数据识别：主动修复错误

引入大模型能力，自动识别数据中的错误内容（如字段缺失、分类错误、格式异常），并主动触发修复流程，实现数据质量的持续自优化。

Affiliations

- 1 University of Rome "Tor Vergata". Italy. RN, MSN,, PhD student. Urdh ri Infermier ve t Shqip ris (UISH) Tirana - ALBANIA Email address: bleriduka@yahoo.it.
- 2 RN, MSN, PhD, FAAN, FFNMRC SI. Centre of Excellence for Nursing Scholarship, Rome, Italy. Email address: alessandro.stievano@gmail.com.
- 3 PT, MSN, PhD student, University of Rome "Tor Vergata". Italy. Urdh ri i Infermier ve t Shqip ris (UISH), Tirana, ALBANIA, Email address: alta_arapi@yahoo.com.
- 4 University of Rome "Tor Vergata". Italy. PT, PhD student. University of Rome "Tor Vergata". Italy. Urdh ri i Infermier ve t Shqip ris (UISH) Tirana - ALBANIA Email address: emanuela.prendi@gmail.com.
- 5 RN, MSN, PhD, FAAN. Catholic University "Our Lady of Good Counsel". Albania. Centre of Excellence for Nursing Scholarship OPI Viale Giulio Cesare, Roma - ITALY Email address: genna.rocco@gmail.com.
- 6 RN, MSN, PhD, FFNMRC SI, Catholic University "Our Lady of Good Counsel". Albania. Centre of Excellence for Nursing Scholarship OPI Roma - ITALY. Email address: ippo66@live.com.

[1]University of Rome "Tor Vergata". Italy. RN, MSN,, PhD student. Urdh ri Infe rmier ve t Shqip ris (UISH) Tirana - ALBANIA.;

[2]RN, MSN, PhD, FAAN, FFNMRC SI. Centre of Excellence for Nursing Scholars hip, Rome, Italy.;

AI识别

[1]University of Rome "Tor Vergata". Italy.;

[2]Centre of Excellence for Nursing Scholarship, Rome, Italy.;

9: 数据加工: 健全的质量与安全监督

质量保障

数据完整性、准确性、及时性

- 数据完整性监控: 核心字段定期检测, 关键字段缺失预警
- 准确性同步验证: 源端与治理后数据对比校验, 异常自动标记
- 加工质量监控: 治理任务执行状态全程可观测, 质量指标可视化

安全合规

数据可靠性、安全性、可控性

- 数据加密与脱敏: 传输与存储全链路加密, 敏感字段自动脱敏处理
- 非法数据过滤: 自动识别并拦截不合规数据进入治理流程
- 分角色权限控制: 数据治理权限与数据利用权限按角色/部门/场景精细授权



第三章

治理成效 数据应用场景

2026年中国高校图书馆数据治理与应用创新研讨会



1. 基于指标体系支持管理决策

指标中心

统一数据度量标准,面向管理决策建立数据模型

- 500+指标库管理
- 60+多维度分析模型
- 指标数据可视化



指标统计标准

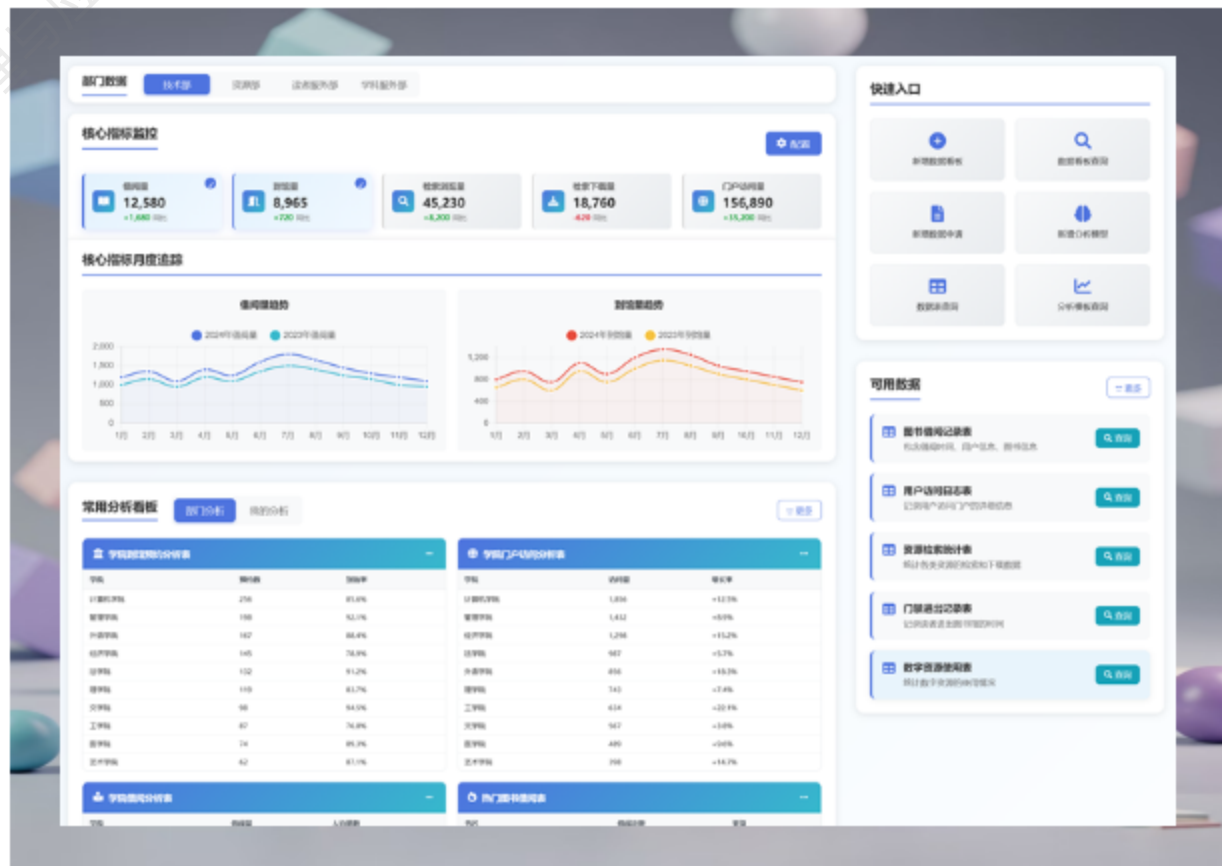
指标类型: [v] 指标名称: [v] 请输入内容 [v] 查找

序号	指标名称	指标类型	数据域	数据源
1	入馆量	派生指标	行为数据域	门禁入馆
2	每月入馆量	派生指标	行为数据域	门禁入馆
3	每月学院入馆量	派生指标	行为数据域	门禁入馆
4	每月读者入馆量	派生指标	行为数据域	门禁入馆
5	读者入馆量	派生指标	行为数据域	门禁入馆
6	年度读者入馆量	派生指标	行为数据域	门禁入馆
7	年度读者场馆入馆量	派生指标	行为数据域	门禁入馆
8	年度读者时段入馆量	派生指标	行为数据域	门禁入馆

数据空间

面向图书馆各部门与读者提供数据服务

- 部门看板: 为各部门提供业务数据看板与分析工具
- 领导看板: 为馆领导提供全局数据与决策建议
- 个人空间: 为读者提供在校期间全部数据集合与报告



2. 基于指标体系的智能套件



馆藏绩效评价

从资产统计、重要学科核心资源保障率、本校成果收录与参考文献保障率提供资源建设有效性的决策数据

- 多维度馆藏分析
- 学科资源覆盖评估
- 成果效益对比分析



阅读报告

面向读者、学院、毕业生提供图书馆服务数据报告与兴趣画像。让读者通过数据了解图书馆服务

- 个人阅读轨迹
- 学院阅读分析
- 毕业学习报告



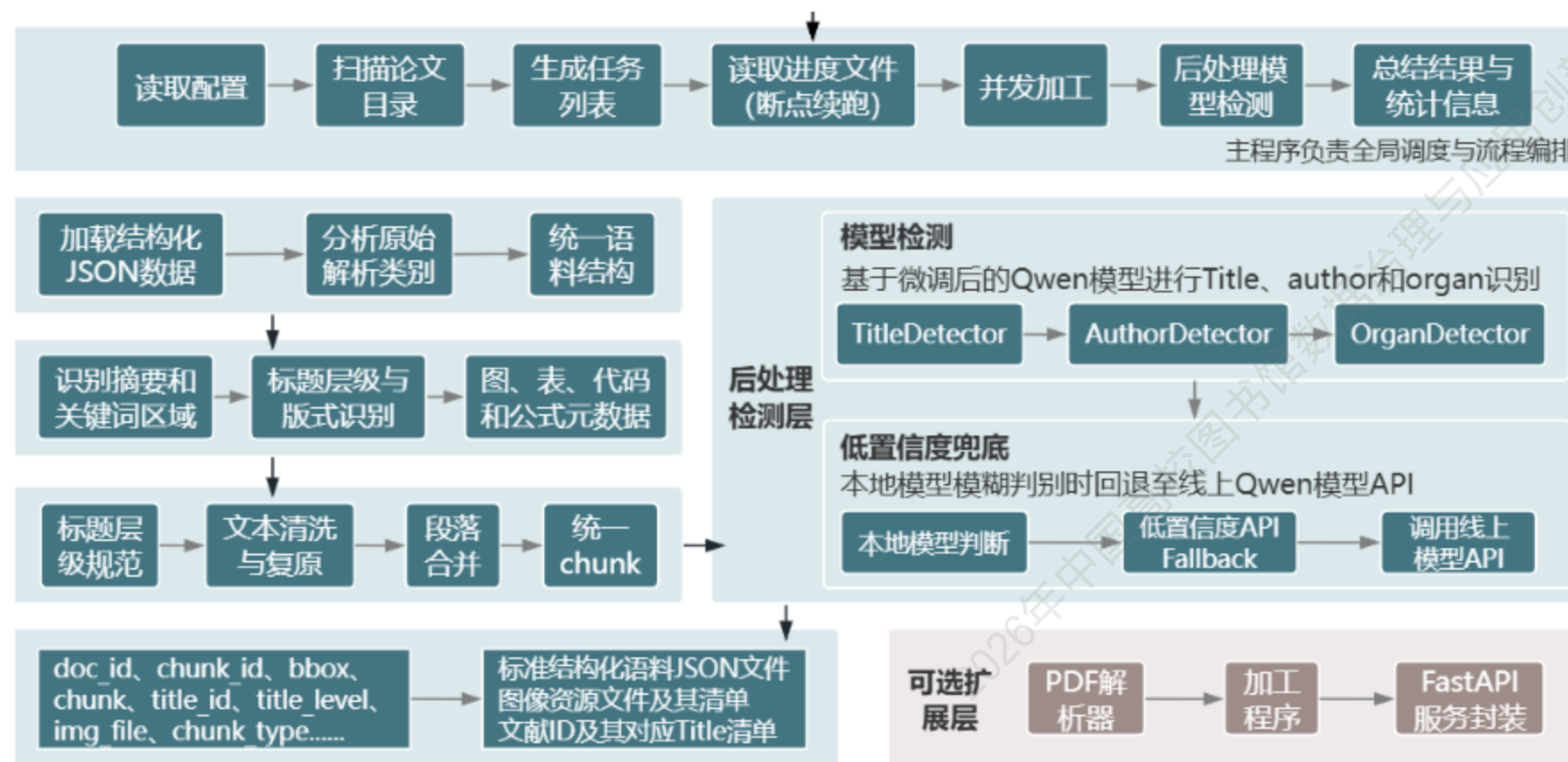
数据驾驶舱 (看板)

数字化孪生图书馆。通过大屏、监视器等终端让读者、领导、馆员即时掌握图书馆当前服务状态

- 实时运行监控
- 关键指标展示
- 异常情况预警

3. 知识语料是高质量数据集的全新服务形态

高质量数据集是驱动面向应用场景、垂直领域服务的核心原料。



语料加工流程

基础语料:

1 FHDW 柴油加氢降凝技术

FHDW 加氢降凝工艺技术使用加氢精制和临氢降凝催化剂组合, 采用加氢精制 - 临氢降凝一段串联工艺原料先进入加氢精制段, 在适宜的反应条件和加氢精制催化剂作用下, 原料进行脱硫、脱氮及芳烃反应, 精制段流出物进入临氢降凝段, 在适宜的工艺条件和临氢降凝催化剂作用下进行择形裂解或反应产物经分离系统与循环氢分离, 然后进入分馏塔分馏出汽、柴油及气体馏分。

1.1 柴油加氢降凝技术原理

柴油的低温流动性与其化学组成有关, 其中碳数较高的正构烷烃的含量是影响柴油凝点的主要因素, 凝点, 就必须降低柴油中正构烷烃含量。在油品构成中高凝点正构烷烃的分子直径最小, 而低凝点芳多侧链的异构烷烃的分子直径较大, FHDW 柴油加氢降凝工艺中所用临氢降凝催化剂, 其作用就

特征语料:

1. 柴油低温流动性核心知识

- 柴油凝点由高碳数正构烷烃含量决定, 正构烷烃分子直径小、凝点高
- 降凝路径: 选择性脱除 / 异构化高凝点正构烷烃, 改变分子结构

2. 加氢降凝核心原理

- 临氢降凝: 分子直径择形催化, 裂解 / 异构高凝点直链烃
- 加氢改质: 多环芳烃加氢开环不断链, 提升十六烷值, 同步脱硫脱氮
- 异构降凝: 正构烷烃异构化, 降凝同时保收率、减密度损失

知识语料:

4. FHDW 技术

- 工艺: 加氢精制 - 临氢降凝一段串联, 双催化剂组合
- 工业化: 90 年代, 6 套装置
- 操作参数: 氢分压 3.4-6.5MPa, 高分压力 3.4-6.7MPa
- 原料: 直馏柴油、焦化柴油、催化柴油混合油
- 产品: -35# 柴油, 凝点 -41~-45°C, 硫≤30ppm, 满足国 IV/ 国 V

4. 基于数据语料升级知识服务



优质文献知识语料



AI检索、AI问答

- 基于语义、实体、关系的智能检索
- 快速获取核心结论、研究方法等
- 帮助用户“精准定位”馆藏资源

高质量数据治理 → 语料知识库 → 智慧知识服务

- 提高用户的知识获取效率、保障知识专业性
- 提高馆藏资源的利用率，提供可信任依据



台风形成的主要条件是什么

台风形成的主要条件包括充足的热量和湿度、海面温度达到26.5°C以上、低层辐合和高层辐散形成的上升运动、以及较弱的风垂直切变等。以下是对这些条件的详细分析：

海温条件

热带气旋（台风）的形成需要海面温度（SST）至少达到26.5°C以上，这是提供足够的热量和湿度的基础条件。温暖的海水通过蒸发作用为大气提供水汽，水汽凝结释放潜热，进一步驱动对流活动 [16]。此外，海温异常（如厄尔尼诺事件）会影响台风生成的频数和位置。例如东太平洋海温升高可能导致西北太平洋台风生成减少 [24]。

大气条件

低层辐合与高层辐散：低层气旋性环流（如热带扰动）为台风生成提供初始涡度，而高层辐散有助于维持上升运动，促进对流发展 [12]。

垂直风切变：较弱的风垂直切变（通常小于10 m/s）有利于台风结构的维持。强风切变会破坏对流组织的对称性，抑制台风发展 [24]。

中层湿度：中层大气的高湿度（相对湿度>70%）有助于对流活动的持续，避免干空气侵入破坏台风内核 [9]。

动力与热力耦合

WISHE机制（风诱导海表热交换不稳定）：台风通过表面风应力增强海气热量交换，进一步强化对流，形成正反馈循环 [1]。

中层涡旋作用：中层涡旋通过层云过程增强低层辐合，同时其冷心结构增加大气不稳定性，促进深对流爆发 [5]。

其他影响因素

大尺度环流：如副热带高压的位置和强度影响台风的移动路径；ENSO事件通过改变沃克环流抑制台

请输入您学术需求，如：请帮我分析中国古代图书编排原因

知识服务平台

5. 基于数据语料提升科研情报效率



情报服务核心价值

图书馆可为科研团队提供：

- 新兴研究热点实时追踪与预警
- 同行研究进展与竞争态势分析
- 学科交叉机会识别与推荐
- 科研成果影响力量化评估

领域文献知识语料

治理数据对象

- 科研文献：期刊、专利、会议论文
- 科研站点：公众号、学术会议、机构网站
- 科研对象：主题、作者、机构画像



The screenshot displays a research platform interface with the following components:

- Navigation:** Top navigation bar with filters: 按研究主题分类 / 按课题组分类 / 按机构分类 / 按期刊分类. A search bar on the right contains "查看全部".
- Filter Tags:** A row of filter tags including C-H Activation, Photochemistry, Electrochemistry, Metal catalyzed reaction, 天然产物, 炎症与免疫调节, 细胞信号转导与基因调控, and 纳米药物递送.
- Article List:** A list of search results with titles like "Amines tuned controllable carbonylation for the synthesis of γ -lactones and 1,4-diones" and "Iron(III)-catalyzed activation of C(sp²)-H arenes".
- Article Detail View:** A detailed view of an article titled "Accelerating Machine Learning Systems via Category Theory: Applications to Spherical Atoms". It includes a "Research News" section, a "本期概述" (This Issue Summary) section, and a "研究进展" (Research Progress) section with sub-points like "1. 壳层衍生磁性材料用于污水深度净化磷酸盐超导体机制突破".
- Reference List:** A "参考资料" (References) section on the right side of the article view.

图书馆从被动响应转变为主动情报供给，深度嵌入科研全流程。

科研服务平台

6. 基于数据语料支撑个人知识管理



用户行为数据语料

通过将读者行为数据与学院、专业、学历、研究方向等标签深度关联，图书馆可精准识别用户的知识需求。

→ 画像构建

识别不同读者的阅读偏好与资源需求

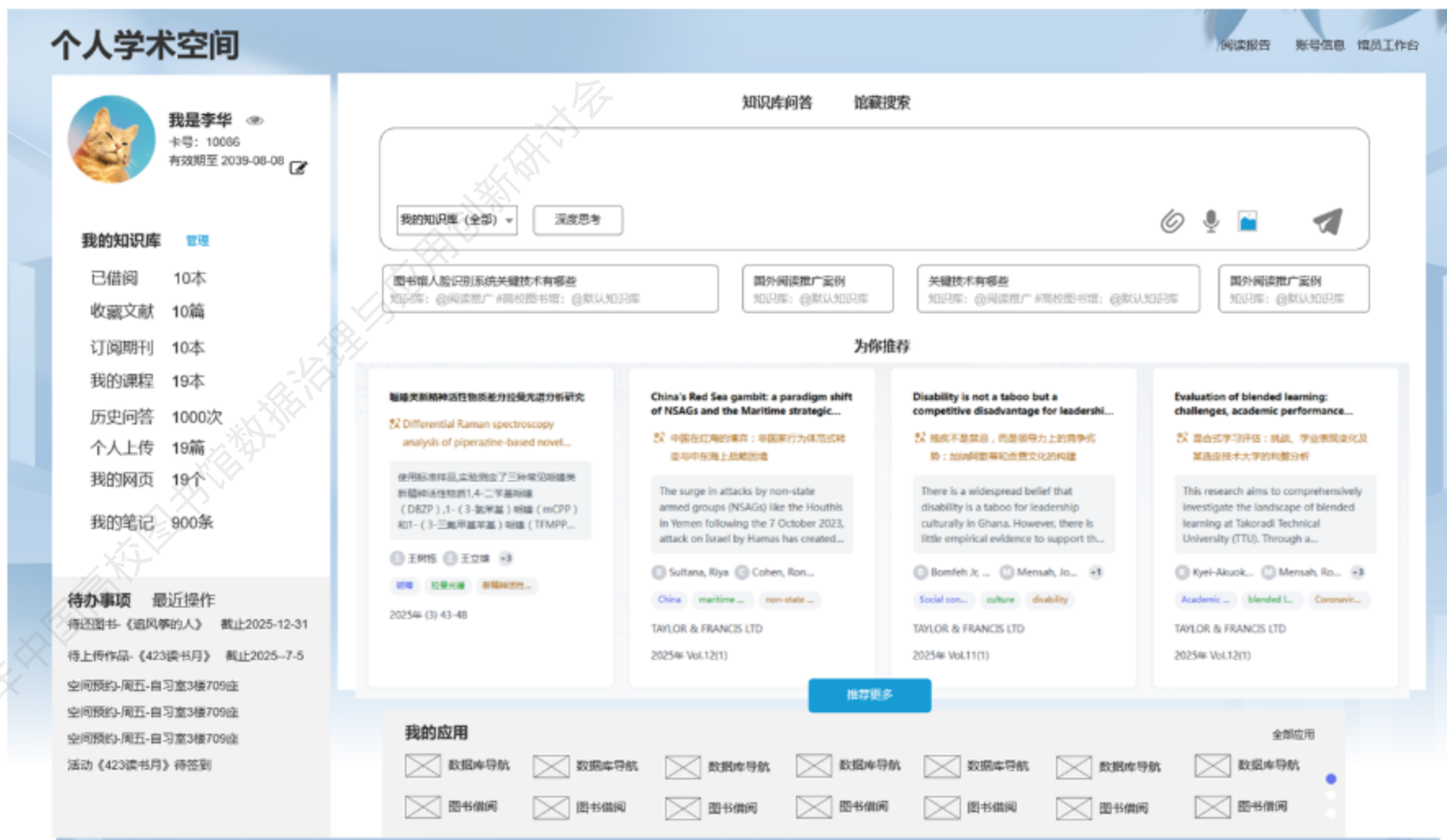
→ 个性化推荐

基于行为数据主动推送相关资源与知识点

→ 学习轨迹追踪

记录读者知识积累历程，初始化个人知识库

体系化的管理知识结构，扩展知识脉络，辅助日常研究学习。



The screenshot shows a user's personal academic space. On the left, a profile for '我是李华' (I am Li Hua) is shown with a cat avatar, ID 10086, and an expiration date of 2039-08-08. Below the profile is a '我的知识库' (My Knowledge Base) section with statistics: 已借阅 10本, 收藏文献 10篇, 订阅期刊 10本, 我的课程 19本, 历史问答 1000次, 个人上传 19篇, 我的网页 19个, 我的笔记 900条. A '待办事项' (To-do list) section follows with items like '待还图书-《追风筝的人》' and '待上传作品-《423读书月》'. The main area features a search bar, a '知识库问答' (Knowledge Base Q&A) section with a search bar and filters, and a '为你推荐' (Recommended for you) section with four article cards. At the bottom, there is a '我的应用' (My Applications) section with a grid of icons for various library services.

个人知识库

7. 通过数据开放扩展服务能力



DALIB 3.0 通过数据开放平台 提供了API、数据总线、MCP接口等多种协议，实现图书馆数据的全面开放与生态合作。释放数据价值最大化。

第三方厂商

支持数据库商、学术平台等第三方系统通过标准接口调用图书馆数据，促进产业协作。

智慧校园

与校级数据中台、教务系统、科研管理平台无缝对接，推动全校数据互联互通。

读者智能体

支持读者个人AI助手通过MCP协议安全调用图书馆数据，实现个性化智能服务。

8. 数据体系建设将重塑馆员分工

新型数据体系落地的关键不只在技术，更在组织与人：数据标准、质量、权限与数据产品需要明确责任人，数据馆员将成为图书馆的新兴核心岗位。

岗位	职责转变	所需新技能
数据馆员（运维向）	账号/服务器运维 → ETL任务监控、数据服务发布、容量管理	数据集成基础、任务调度、API网关
数据馆员（编目向）	编目与资源维护 → 标准模板、分类映射、字段口径统一	元数据标准、数据字典管理
数据馆员（分析向）	出报表 → 指标口径治理、数据分析设计、服务化交付	指标体系、数据建模思维
读者服务馆员	咨询/活动 → 基于数据的服务改进、读者分群与服务评估	数据分析素养、A/B思维、隐私合规
信息化馆员	网络安全 → 数据分级分类、脱敏策略、合规审计	数据安全法/PIPL理解、审计演练
学科服务馆员	导航与培训 → RAG 语料策划、AI知识库运营、效果评估	RAG与知识库运营、提示词规范

9. 数据馆员能力的系统性重构



数据馆员培训计划

- 基础培训**
数据治理理念与DALIB平台操作
- 进阶培训**
数据配置、数据建模、指标设计
- 高级培训**
数据分析、报告撰写、数据创新
- 案例研讨**
标杆案例学习与实践经验交流

DALIB提供完整的培训体系与持续的技术支持,确保馆员快速掌握数据管理技能

10. 数据治理成功的标志：赋能服务

衡量图书馆的数据化转型是否真正完成，需要经历三个递进阶段：

数据资源

原始、零散的数据，尚未经过系统治理，仅作为运营副产品存在

数据资产

经过治理与标准化的数据，能够有效支撑业务使用，具备可管理、可度量的资产属性

数据要素

数据作为产品直接进入服务市场，产生经济属性与知识价值，成为图书馆核心竞争力

☐ DALIB 是帮助图书馆完成从数据资源到数据要素完整跨越的最优平台，提供全阶段的技术支撑与方法论分享。

可信数据基座

DALIB建设可信、可控、可追溯的数据基础设施，确保每条数据都有规范的处理过程与可验证的质量记录，为服务提供坚实支撑。

业务场景赋能工具

DALIB是贴合业务场景的赋能工具——数据治理目标对应真实的馆务痛点与读者需求，让数据真正从“存储的资产”转变为“流动的服务”。

11. 数据已经成为众多核心场景的重要支撑



图书馆的数据基座已经面向应用、智能体等形态在知识问答、科研情报、教学辅助、馆务决策等场景提供核心支撑。

第四章

数据价值展望

数据价值未来的发展路径，已经清晰可见

2026年中国高校图书馆数据治理与应用创新研讨会



1.业务成长关键：高质量数据语料服务

基于数据的深度知识内容加工，构建高质量知识语料，将成为图书馆面向高校核心目标的重要能力支柱。

语料服务的战略价值

AI和数据会加快重构教学、科研和管理全流程，图书馆将不再是资源的被动存放地，语料服务将成为图书馆面向高校核心目标的重要能力。

DALIB 2026 数据语料服务

DALIB 将于 2026年正式推出数据语料服务模块，助力图书馆搭建数据门户，面向高校师生与科研团队提供：

- 垂直学科高质量语料包
- 结构化知识实体数据集
- 语料订阅与按需调用服务



数据图书馆
文献数据浏览与检索

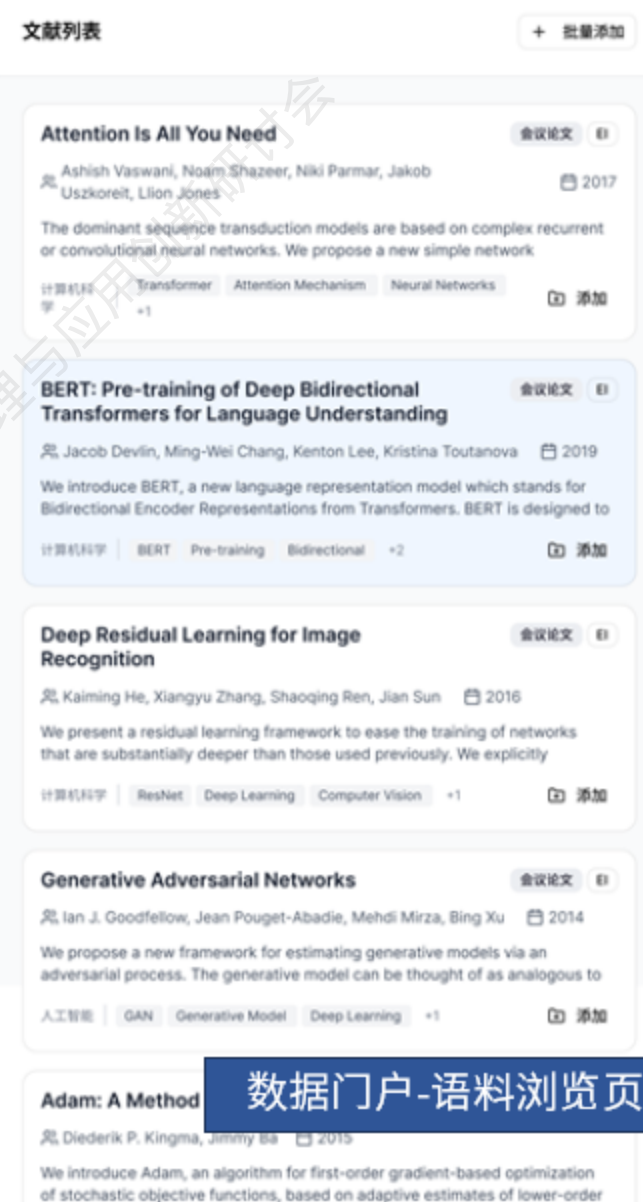
自定义主题检索
标题、关键词、作者...
支持按标题、关键词、作者名称搜索

文献类型
 期刊论文
 会议论文
 学位论文
 专利
 技术报告

学科分类
 计算机科学
 密码学
 医学信息学
 生物信息学
 人工智能
 数据科学

核心收录
 SCI
 EI
 CSSCI
 CSCD

共找到 20 条文献



文献列表 + 批量添加

Attention is All You Need 会议论文 EI
Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones 2017
The dominant sequence transduction models are based on complex recurrent or convolutional neural networks. We propose a new simple network
计算机科学 | Transformer | Attention Mechanism | Neural Networks +1 添加

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding 会议论文 EI
Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova 2019
We introduce BERT, a new language representation model which stands for Bidirectional Encoder Representations from Transformers. BERT is designed to
计算机科学 | BERT | Pre-training | Bidirectional +2 添加

Deep Residual Learning for Image Recognition 会议论文 EI
Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun 2016
We present a residual learning framework to ease the training of networks that are substantially deeper than those used previously. We explicitly
计算机科学 | ResNet | Deep Learning | Computer Vision +1 添加

Generative Adversarial Networks 会议论文 EI
Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu 2014
We propose a new framework for estimating generative models via an adversarial process. The generative model can be thought of as analogous to
人工智能 | GAN | Generative Model | Deep Learning +1 添加

Adam: A Method
Diederik P. Kingma, Jimmy Ba 2015
We introduce Adam, an algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order



BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
会议论文 EI 计算机科学

基础数据 特征语料 知识语料

三元组信息

- BERT 基于 Transformer编码器
- BERT-Base 层数 12层
- BERT-Large 层数 24层
- 预训练 任务 掩码语言模型
- 预训练 任务 下一句预测
- SQuAD 1.1 F1分数 93.2
- 词表 大小 30000

知识图谱



数据门户-语料浏览页

2.迈向"十五五"数据智慧新生态

数据治理是图书馆服务框架重构的起点，是通往"智慧中枢"的必经之路。
面向十五五，通过**数据语料+AI智能体**，助力图书馆深度融入学习、教学、
科研等需求场景



2026年中国高校图书馆数据治理与应用创新研讨会



数智融合 赋能未来

中国高校图书馆数据治理与应用创新研讨会

谢谢

数据成就智慧 知识赋能未来



维普智图

VIPSMART

2026年中国高校图书馆数据治理与应用创新研讨会