



项目编号：2020054

CALIS 全国农学文献信息中心研究项目

基于知识图谱的高校智慧图书馆建设

研
究
报
告



仲恺农业工程学院

2021 年 4 月

目录

| | |
|---------------------|----|
| 1 研究概述 | 1 |
| 1.1 研究背景 | 1 |
| 1.2 研究目的及意义 | 3 |
| 1.3 研究内容 | 3 |
| 1.4 建设目标 | 4 |
| 2 项目实施方案 | 5 |
| 2.1 研究团队组成及分工 | 5 |
| 2.1 研究思路 | 8 |
| 2.2 技术路线 | 9 |
| 2.3 研究方法 | 10 |
| 2.4 研究内容 | 10 |
| 3 项目特色及价值 | 14 |
| 4 成果形式 | 15 |

基于知识图谱的高校智慧图书馆建设 研究报告

1 研究概述

1.1 研究背景

2009年，“智慧地球”概念一经提出，即在全球范围内引起关注。中国作为世界上最大的发展中国家，“中国智慧”成为近几年最热门的词汇之一，全国范围内的“智慧城市”建设也陆续开展起来。芬兰奥卢大学图书馆学者 Aittola 提出“智慧图书馆”将成为智慧城市建设的重要内容，他认为：智慧图书馆是基于位置感知的，能够帮助用户查找所需文献、资料，是一种移动图书馆服务。智慧图书馆作为未来图书馆的发展方向，是“智慧城市”公共文化服务体系中的重要组成部分之一，其理念研究和服务方式探索成为图书馆学者们近年来持续关注的焦点和热点。智慧图书馆服务的宗旨是为广大读者提供更加方便快捷的信息查询和更加智慧化的综合服务，可以让读者随时随地享受图书馆资源，满足不同读者的阅读需求。2017年，田长斌认为智慧图书馆应是以物联网、RFID、云计算、大数据、移动通信等技术为基础，以智能手机、智能芯片、无线传感等智慧化设备为手

段，使海量馆藏资源实现书书相联、书人相联、人人相联，为读者提供全方位、一体化的智慧图书管理服务模式。当前，智慧图书馆已成为未来图书馆创新发展的新方向。

近年来，随着计算机技术、可视化技术的迅猛发展及其在科学领域的广泛应用，使得人们获取知识和处理知识的能力大大加强，一种以可视化的图像直观展现出来的新兴交叉学科——知识图谱悄然兴起，这是科学计量学、文献计量学的新发展。知识图谱(Knowledge Graph)，也称为科学知识图谱、知识域可视化或知识域映射地图，是显示科学知识的发展进程与结构关系的一系列各种不同的图形，它通过将应用数学、图形学、信息可视化技术、信息科学等学科的理论与方法与计量学引文分析、共现分析等方法结合，并利用可视化的图谱形象地展示学科的核心结构、发展历史、前沿领域以及整体知识架构达到多学科融合目的，为学科研究提供切实的、有价值的参考。勾勒知识与知识之间的相互联系，把复杂的科学知识通过图形的形式呈现出来，从而使人们以更直观的形式了解知识，了解该学科及领域的知识研究现状，为今后的知识研究、学术探讨提供科学依据。

知识图谱在信息技术、可视化技术及文献计量学等多种学科的交叉中产生，提供了理顺知识脉络的可视化体系网络，借以了解某个知识领域的动态、热点，预测和把握知识发展的前沿。知识图谱能够打破不同领域、不同场景下的数据隔离，聚合信息和知识，为智慧性的搜索、推荐、问答、解释与决策等应用提供基础支撑。知识图谱对高校图书馆实现知识化、个性化、场景化、多元化的智慧服务有独特优

势，使“知识之网”愿景成为了可能。从已有研究来看，知识图谱可以可视化地揭示领域研究前沿及发展趋势，具有良好的交互性和傻瓜式的图示形式，越来越受青睐，并被应用于多个学科、多个领域的研究中。与此同时，在信息技术高速发展、知识老化速度加快的背景下，无论是公共图书馆还是高校图书馆，其服务方式和服务内容，都需要作出相应的改变，以便更加贴切地满足读者的需求。

1.2 研究目的及意义

将知识图谱技术运用到图书馆建设中，创新高校图书馆智慧检索、智慧推荐、智慧问答等智慧服务模式，实现用户、馆员、馆藏资源互联互通的智慧化图书馆，不仅可以满足用户个性化、多元化和主动性的信息服务需求，而且能让读者快速、有效从海量的文献资源中搜索到高质量的信息，快速提升图书馆的服务及管理水平，为高校教学、科研、学科建设、管理决策提供全面支持。

1.3 研究内容

(1) 聚合和挖掘海量馆藏资源

高校图书馆文献资源种类繁多、专业性强、载体形式多样，对文献资源建立文献知识图谱，通过挖掘文献提炼精品知识建立学科知识图谱，分析用户大数据建立用户图谱。以知识图谱的形式对各类信息和知识进行聚集、融合和组织，形成有内在逻辑关联、对智慧服务有

意义的知识结构体。在此基础上，开展计算和应用，将减轻用户的认知负荷，帮助用户摆脱信息过载和知识迷航的困境。

(2) 研究融合和利用多源用户数据

移动互联网、社交网络、可穿戴设备的应用使图书馆形成了多源异构、实时变化的海量用户数据。图书馆用户数据包含用户的背景信息、借阅记录以及点击、下载、浏览、收藏、评论等用户行为日志。此外，在图书馆的微博、微信、论坛、QQ等服务社区中，用户和图书馆的互动和交流产生了社交数据。融合多源用户数据以认知和表达用户需求是智慧服务面临的挑战。构建用户图谱，全景式地描述用户的兴趣、需求、特征以及用户之间的多样、动态关联，以理解用户和洞察用户。

(3) 探究感知和识别多样动态场景

场景服务需要从用户的角度出发，捕捉用户在特定场景下的需求。场景具有动态性和多样性，随用户的行为不断切换。感知和识别多样动态场景以提供场景服务是智慧服务亟待解决的问题。场景要素包括用户、目的、行为、时间、地点等，在场景、场景要素之间建立联系，以场景知识图谱的形式组织和存储多样场景的知识。通过场景知识图谱，结合用户的行为、时间、地点等，以便感应用户所处的场景。

1.4 建设目标

- (1) 发表研究文论 1 篇；
- (2) 提交研究报告 1 份；

(3) 申请软件著作权 1 个。

2 项目实施方案

2.1 研究团队组成及分工

(1) 研究团队情况

研究团队由计算机技术、网络技术、信息管理、应用数学、图书资料、档案管理等专业的专家和技术骨干组成，长期活跃在教学和科研第一线，主要从事信息系统设计与开发、物联网、云计算与大数据、智能信息处理与挖掘、数学建模等教学科研工作，拥有丰富的实践经验，具有较强的教学及科研能力。团队成员全部具有硕士以上学位，其中，教授 3 人、副教授 3 人、中级 6 人、初级 2 人，平均年龄约 38 岁，学历、职称和年龄结构合理，形成了一支年富力强的教学科研团队。团队成员近五年来主持国家级、省部级、市厅级等各类教学科研项目 60 多项，授权专利 50 余件，软件著作权 40 多个，公开发表论文 60 余篇、获得各类奖项 30 多项，主编及参编专著 10 余部。同时依托校内外创业孵化基地，采用形式多样的方式培养研究生的创新能力，近 5 年指导学生获得各类竞赛奖 20 多项，获得知识产权 10 余件，创新创业训练计划项目立项 10 多项，科技创新培育专项资项目（攀登计划）立项近 10 项。

(2) 研究团队分工

| 姓名 | 专业技术职务 | 项目中的分工 |
|-----|--------|------------|
| 李湘丽 | 高级工程师 | 项目负责人 |
| 曹亮 | 工程师 | 项目实施 |
| 黄娟 | 助理馆员 | 材料撰写 |
| 陈志慧 | 研究馆员 | 实地考察、调研 |
| 田思 | 助理馆员 | 资料收集、整理 |
| 郭鹏飞 | 讲师 | 数据建模 |
| 徐浩根 | 工程师 | 系统开发 |
| 卢娜 | 高级工程师 | 数据库构建 |
| 呼增 | 讲师 | 无线网络优化 |
| 杨雁 | 工程师 | 数据数字化处理及加工 |
| 刘双印 | 教授 | 项目指导、协调 |
| 徐龙琴 | 教授 | 平台设计与指导 |
| 贺超波 | 副教授 | 数据挖掘、数据构建 |
| 王潇 | 讲师 | 数据分析 |

(3) 团队依托平台

团队依托“仲恺农业工程学院图书馆”、“广东省农产品安全大数据工程技术研究中心”、“广东高校智慧农业工程技术研究中心、智慧农业创新研究院等多个省级科技创新平台和广东省级计算机实验教学示范中心、农业物联网虚拟仿真实验教学示范中心、电工电

子实验教学示范中心等 3 个省级教学示范中心，一直在关注和跟踪现代图书馆信息建设的发展动态，已进行了大量资料调研和前期准备工作，为课题的研究积累了丰富素材，奠定了较好的基础。

(4) 实验设备条件

学校图书馆馆藏图书总量达到 354.43 万册，其中：纸本图书 122.43 万册，电子图书 232 万册；中外文纸质期刊 350 种，电子期刊 43639 种；音频资源 10818 小时，中外文电子数据库 43 个。文献资源覆盖农、工、文、理、法、经、管、艺术等多学科，已初步形成了多学科、多载体、以热带、亚热带农业科学、生物科学文献为特色和优势的文献保障体系。图书馆设有办公室、流通部、信息咨询部、技术部、采编部，拥有 8 个集藏、借、阅、检索等服务功能为一体的书刊阅览室，可提供阅览座位 1800 多个，电子阅览机位 170 个。图书馆拥有大数据资源服务器，网络资源实行 24×7 不间断服务，业务工作全部实现计算机管理，网络环境良好，软硬件设施较为完备。

(5) 保障措施

研究团队已构建了仲恺农业工程学院图书馆网站管理系统、基于数据挖掘的专利共享信息服务平台、基于物联网的监控云服务平台、物联网 APP 等相关平台 10 多个，获得“科技文献合著关系网络主题社区挖掘软件”、“仲恺农业工程学院图书馆网站管理系统”、“基于语义本体树的科研公告协同推送平台”、“基于大数据挖掘的网络舆情智能分析系统”、“基于非负矩阵分解的社交网络好友推荐软件”、“大数据专利信息共享系统”、“基于蚁群算法的应急物资调度系统”、

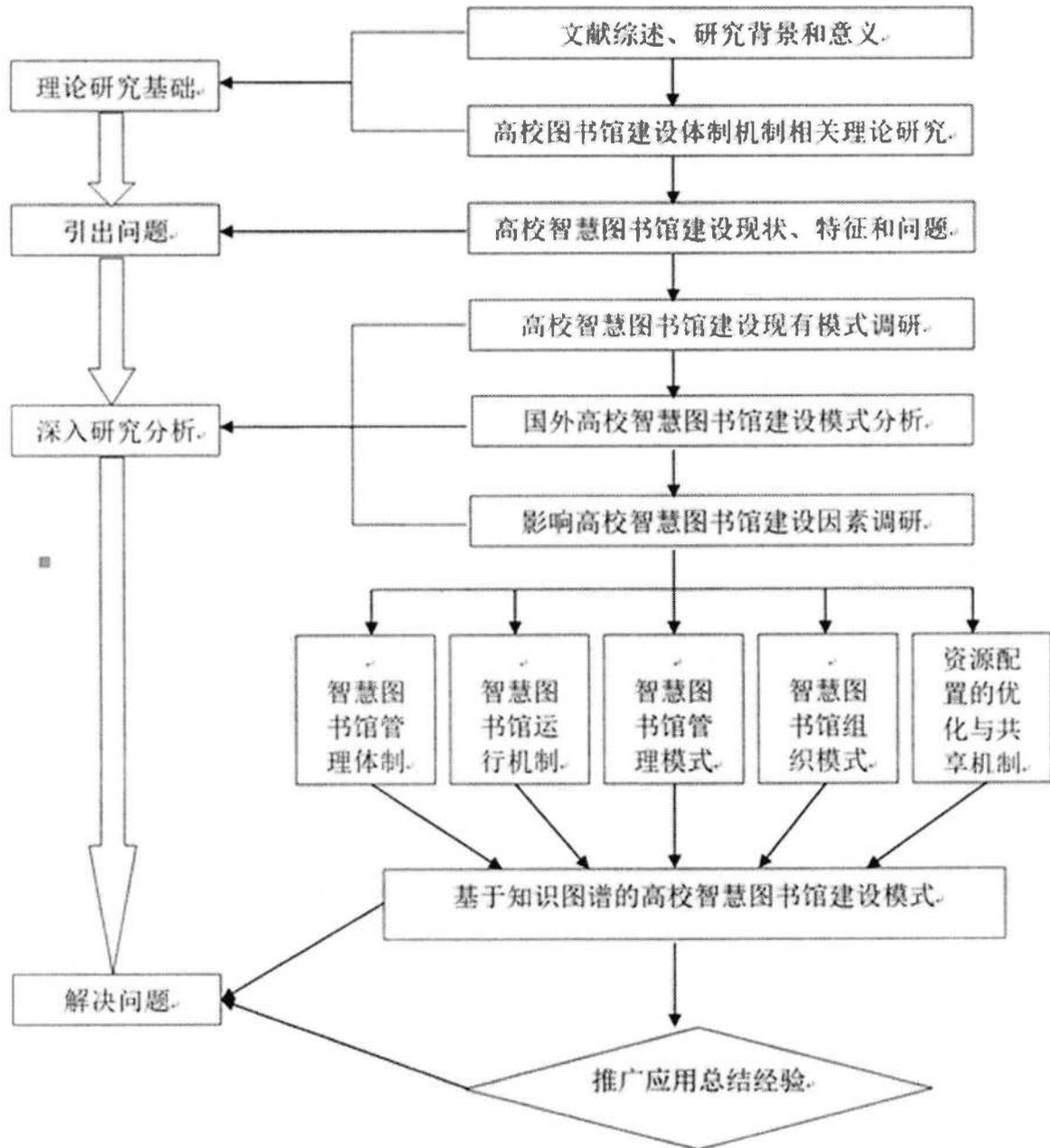
“基于人工智能和大数据的农业物联网故障诊断系统”等软件著作权，为本项目提供了丰富的研究资料和必要的人力、物力和财力的条件保障。

项目承担单位仲恺农业工程学院为项目研究提供科研条件和师资力量，并且学校一直十分重视科研课题的管理工作，给予各课题组大力的支持和配合，为各课题提供必要的研究条件和保证足够的研究时间，对课题给予资助经费以确保课题的顺利完成，同时学校具有良好的财务纪律，长期以来严格执行《仲恺农业工程学院财务管理办法》和《仲恺农业工程学院科技项目及经费管理办法》等，可保证科研经费专款专用，能提供确保完成本项目全部研究任务的各项保障措施。

2.1 研究思路

主要通过“理论基础研究”→“引出问题”→“深入研究分析”→“解决问题”的思路进行项目研究。

2.2 技术路线



2.3 研究方法

(1) 文献调查法：通过文献检索，全面了解国内外智慧图书馆发展动态及研究成果，以便更有针对性地做好本研究。

(2) 实地调研：调研兄弟院校图书馆信息化建设情况，分析建设成效和经验，研究智慧图书馆建设技术集成方法。

(3) 实验研究与方案设计：提出基于知识图谱的可动态扩展的智慧图书馆构建模型，进行实验平台部署与初步应用。

(4) 软件设计与开发：进行基于知识图谱的智慧图书馆设计与模型开发。

2.4 研究内容

(1) 高校智慧图书馆中的知识抽取

在高校智慧图书馆馆藏资源中，文献名和作者之间就是显式关系，同一作者的文献名之间的关系就是隐式关系，关系抽取采用马尔可夫逻辑网联合推理关系抽取方法。属性抽取即把实体本身的属性、实体属性与属性之间、开放域数据集中属性进行提取，形成属性对应关系，再进行抽取。高校智慧图书馆中知识抽取如图 1。

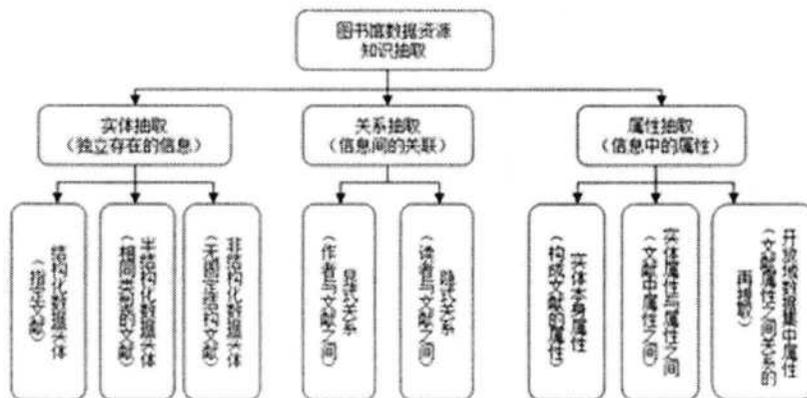


图 1 高校智慧图书馆知识抽取关系图

基于知识图谱的智慧图书馆就是从不同来源、不同结构的数据中进行抽取，形成数据知识存入到知识图谱库中。

(2) 高校智慧图书馆中的知识表示

高校智慧图书馆中，实体之间的关系、关系之间的关系存在着复杂的交叉关系，例如：作者之间、读者之间、文章之间、作者和文章之间、读者和文章之间的关系等，都存在着交叉关系，如图 2。

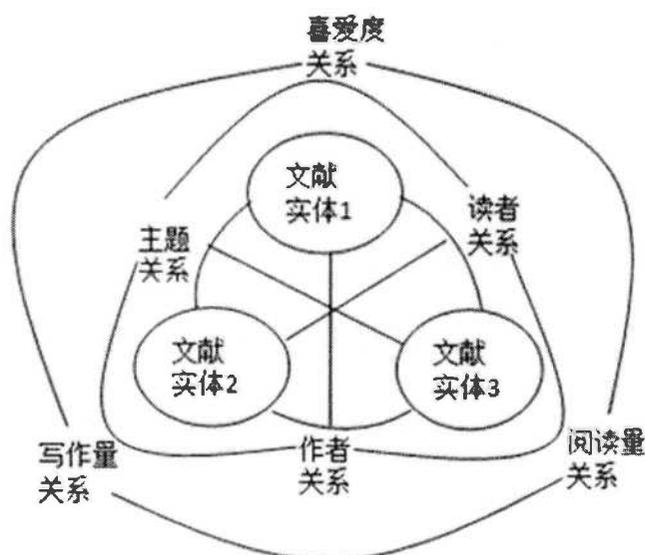


图 2 高校智慧图书馆实体关系图

通过知识图谱的知识表示,就可有效把高校智慧图书馆多源信息表示出来,给人以清晰和明确的认识和理解。

(3) 高校智慧图书馆中的知识融合

知识图谱中的知识融合可让高校智慧图书馆中多样性、多源性和重复性的知识数据有效聚合。通过在同一规范下进行实体对齐、实体链接和关系推演等操作,把不同来源的知识进行融合。知识融合在高校智慧图书馆中的技术流程如图 3: 首先对目标读者和所需数据进行预处理,剔除无用信息后用 Hash 函数进行分块处理,再把原数据信息和目标数据信息相对应,根据对应的复杂度确定是否进行四层负载均衡处理,然后把原数据信息和目标数据信息进行属性相似度和实体相似度计算,并给出相对应链接,最后对知识融合的结果进行反馈评估参数校正,最终输出正确结果或反馈原因。

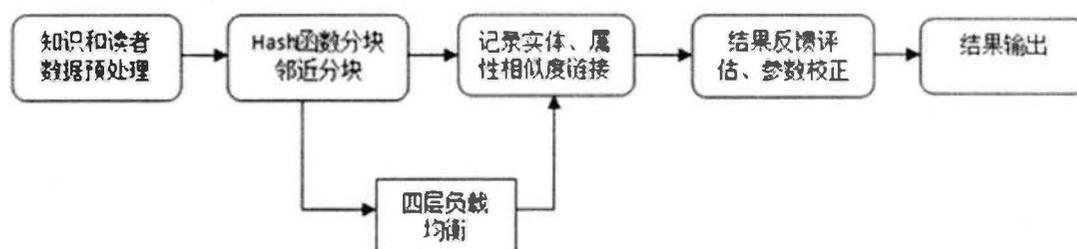


图 3 高校智慧图书馆知识融合技术流程图

通过知识图谱的知识融合将智慧图书馆中数量庞大、种类繁多和关系复杂的资源数据进行深度融合处理,揭示馆藏资源的潜在知识关联规律,为研究数据价值提供参考。

(4) 高校智慧图书馆中的知识推理

高校智慧图书馆中的知识推理就是将馆藏资源实体知识经过推理,挖掘出隐含的关联信息和知识点。其中实体知识包括实体及其属

性，已知或是推理出的实体间关系可以再次作为新实体知识被推理，如此循环可反复挖掘出有价值信息。图4为高校智慧图书馆一个知识推理的具体实例：图中实线为读者1借阅路线，长虚线为读者2借阅路线。通过基于图的知识推理，读者1和读者2都借阅了考研必须的公共课和基础课书籍，应为准备考研的读者；另外通过专业课反推，可知，读者1报考计算机类相关的研究生，读者2报考农业园艺类相关的研究生。

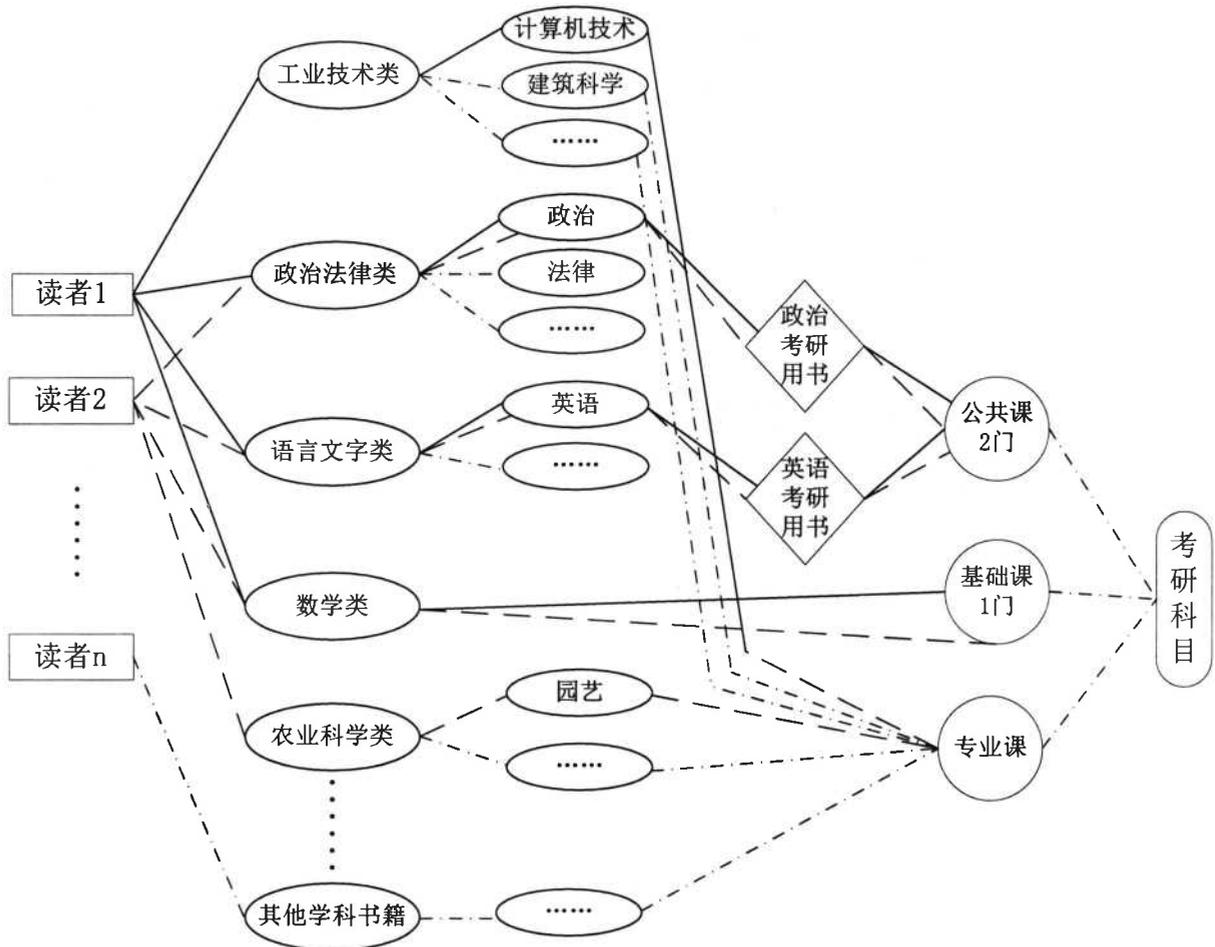


图4 高校智慧图书馆知识推理

高校智慧图书馆海量的、复杂的的藏资资源，采用知识推理可以有效挖掘和推理出潜在的关系信息，通过提供接近“专、精、深”的搜索查询功能，为用户提供“懂需求”的智能服务，达到精准推送。

(5) 知识图谱的智慧荐购

大数据背景下的高校智慧图书馆，利用知识图谱技术能有效将多数读者偏爱的文献资料、热门搜索书籍和高质量馆藏资源进行抽取加工、融合处理，智慧推荐出读者所需的实体书或电子书，快速提高智慧推荐资源的质量和精度。

(6) 知识图谱的智慧荐阅

智慧图书馆通过读者的身份信息、以往搜索信息的关联性、曾阅读的数据信息、相关专业等常用信息等数据信息，运用知识图谱技术进行分析和匹配，快速得出数据信息之间的相关性和关联性，及时将最优结果（最接近读者需求的文献资源）推荐给用户，满足用户最优智慧推荐阅读体验。

3 项目特色及价值

(1) 建立馆藏资源知识图谱。图书馆拥有海量文献，其中蕴含的大量知识是图书馆信息服务和知识服务的重要资源，有效聚合和挖掘海量馆藏资源以提供知识服务是智慧服务的挑战之一，也是本项目的特色之一。

(2) 多源信息融合中的知识表示。在知识融合表示中，融合是最关键的前期步骤，如能有机的融合多源异质的实体、关系等信息，将有利于进一步提升知识表示模型的区分能力以及性能。基于实体的、关系的、Web 文本的、多知识库的融合均具有较为广阔的研究前景，也是本项目的另一个特色。

知识图谱在高校智慧图书馆建设中提供精准的上传下达服务，为知识的大数据化合理利用提供优质路径，为读者与高校智慧图书馆之间建立起了网络纽带。在未来，基于知识图谱的高校智慧图书馆将结合人工智能技术，把书、人、时间和空间有效融合为一个知识生态有机体，通过对读者各种行为数据分析，主动推荐图书、推送资源，构建“会猜想、懂（你）心思”的现代智慧图书馆，实现无障碍、横跨时空的资源共享，资源利用的最大化，为读者带来便捷、生动、有趣和智能的阅读体会。

4 成果形式

(1) 发表研究文论

| 篇名 | 作者 | 出版刊物 | 出版时间 | 备注 |
|--------------------------|--------------------|--------|----------|----------|
| 基于知识图谱的高校智慧图书馆建设研究* | 曹亮，李湘丽*（通讯作者），刘双印等 | 无线互联科技 | 2021年9月 | 已录用，9月出版 |
| 基于Neo4j图数据库的对虾养殖领域知识图谱研究 | 郑祉盈，曹亮（通讯作者），李湘丽等 | 通讯世界 | 2020年11月 | |

(2) 软件著作权

| 成果类型 | 名称 | 登记号 | 著作权人 | 登记日期 | 状态 |
|-------|--------------------|---------------|--------------------|------------|-----|
| 软件著作权 | 基于知识图谱的高校智慧图书馆管理系统 | 2021SR0304123 | 仲恺农业工程学院 李湘丽、曹亮 | 2021.02.26 | 已登记 |

(3) 发明专利

| 成果类型 | 名称 | 专利号 (申请号) | 发明人 | 状态 |
|------|---------------------|----------------|--------------------|--------|
| 发明专利 | 一种基于数据挖掘的专利信息预测系统 | 202011351495.4 | 曹亮, 李湘丽, 刘双印, 徐龙琴等 | 已受理、实审 |
| 发明专利 | 基于知识图谱的农产品智能推荐方法及系统 | 202010465271.X | 曹亮, 李湘丽, 刘双印, 徐龙琴等 | 已受理、实审 |
| 发明专利 | 一种水禽知识图谱构建方法及装置 | 202010347695.6 | 刘双印, 徐龙琴, 李湘丽, 曹亮等 | 已受理、实审 |

(4) 获奖

| 获奖项目 | 获奖等级 | 获奖人 (指导老师) | 获奖时间 | 颁发单位 |
|---------------------------|------|------------|---------|-------------------|
| 2020年粤港澳大湾区IT应用系统开发大赛广州赛区 | 二等奖 | 李湘丽 | 2020.11 | 广州市科学技术协会 |
| 2020年广东省大学生计算机设计大赛 | 一等奖 | 曹亮、徐龙琴 | 2020.09 | 广东省教育厅 |
| 2020年广东省大学生计算机设计大赛优秀指导教师 | 优秀 | 曹亮 | 2020.09 | 中国大学生计算机设计大赛组织委员会 |