



基于用户数据驱动的智能书目系统构建^{*}

彭贤哲 石 进 李 明

[摘 要] 信息资源时代下,作为知识载体的文献资源与日俱增,对书目工作的效率和服务质量提出了新的时代需求。总结前人的研究工作,发现书目的最终导向在于服务用户,但传统书目在实践工作中多以资源为导向,用户的主体地位不够显著。为此,当代书目工作应纳入用户因素,重点在于处理书目资源库、用户需求、用户能力三者之间的关系,由此衍生出智能书目的概念及内涵,据此设计构建的智能书目系统,通过动态建设深度标引组织的书目资源库、针对性满足用户各类个性化需求、精准定位用户各项能力,能够扩展书目资源库、用户需求、用户能力三者之间的交集,提供相宜的灵活智能性服务,推动书目的智能个性化发展进程,助力智慧图书馆知识服务平台建设。

[关键词] 智能书目 书目资源 用户需求 用户能力 智慧图书馆

[中图分类号] G250.7 **[文献标志码]** A **[DOI]** 10.19764/j.cnki.tsgjs.20222005

[本文引用格式] 彭贤哲,石进,李明.基于用户数据驱动的智能书目系统构建[J].图书馆建设,2024(1):108-120.

Construction of Multi-User Data Driven Intelligent Bibliographic System

Peng Xianzhe, Shi Jin, Li Ming

[Abstract] At the era of big data, an increasing number of literatures as knowledge carrier forces the improvement in bibliographic works. Previous research for bibliographic works usually tends to ignore users' needs for potential knowledge. Thus, this study insists that bibliographic works should take users' information into account and focus on the relationship among the bibliographic resources, the needs from users and the ability of users. This viewpoint results in the production of intelligent bibliography. Intelligent bibliography system oriented by users concentrates on the collection and organization for bibliographic resources marked by superficial and profound information, satisfaction for various needs from users, and assessment for users' ability in all sorts of fields. It will provide intelligent services by expanding the intersection of bibliographic resources, needs from users and users' ability. To some extent, it is feasible to improve the intelligent level of bibliographic work and construct the knowledge service platform of smart library.

[Keywords] Intelligent bibliography; Bibliographic resource; User's demand; User's ability; Smart library

0 引 言

书目旨在揭示文献的内外部特征,通过有效提炼、组织书籍的核心知识内容,可以缩短书籍与用户的知识距离,降低用户获取资源的成本。然而长久以来,书目工作多以书籍文献资源的建设为核心,遵循“资源决定服务、服务决定需求”的模式,忽略了用户在书目使用过程中的主体地位,由此构建的书目显然不适宜当下以人为本的服务环境。知识经济时代,书目的工作导向应转向“以用户为中心”^[1],书目构建模式需过渡为“需求决定服务、服务决定资源”^[2],人

工智能、大数据分析等新兴技术的出现为书目工作模式转换提供了现实途径,有利于书目工具在新时代背景下的转型和变革,使其适应“十四五”发展规划中全国智慧图书馆体系的构建,营造与用户共创的开放式知识服务环境^[3]。

为此,本项研究响应智慧图书馆“以人为本、绿色发展、方便读者”的号召^[4],提出了“以用户为中心”的“智能书目”,通过将用户因素纳入书目建设的工作中,由收集的用户数据驱动书目的组织与维护,促成文献资源的智能化治理,推进书目工具的智能化进程,构建不同用户的个性化书目工具,为大数据环境下需求多样、但力有不逮的用户获取文献资源,提供

^{*} 本文系国家社会科学基金项目“面向国家安全的科技情报态势感知研究”的成果之一,项目编号:21BTQ012。



一个知其所缺、晓其所长、明其所需、予其所求的专属智能助手工具,推动智慧图书馆中知识服务平台个性化功能的构建与完善。

1 智能书目研究基础

1.1 相关研究

书目作为文献资源管理利用的工具,一直以来都是图书馆系统迭代优化工作的重心。自1950年起,图书馆特定业务独立自动化系统的开发,既已涉及书目编制等诸多任务工作^[5]。进入1980年,MARC作为书目元数据格式的出现与发展^[6],促成了图书馆系统由钻研单一任务自动化业务^[7]向集成独立功能模块^[8]的转型。之后,元数据领域FRBR (Functional Requirements for Bibliographic Records,书目记录的功能需求)、RDA (Resource Description and Access,资源描述与检索)、BIBFRAME (Bibliographic Framework Initiative,书目框架)等基于关联数据设计书目本体方案,陆续应用于图书馆知识组织与融合。书目编制方案的多样性虽然有利于文献资源描述的全面性,但却阻碍了资源的统一管理利用。因此,下一代图书馆管理系统的研究方向,在于利用云计算、Web技术和发现系统,支持各种开放元数据格式和数据交换协议,实现跨媒体的书目资源描述与组织^[5, 9],如FOLIO (The Future of Library is Open,云瀚)平台开发项目的元数据资源集成中心Codex^[10-11]。

传统意义上书目的功能主要是为用户提供文献信息、提高文献查找速度和效率、最大限度地满足用户对文献信息的特定需求^[12]。参照图书馆系统演化先后经历了以纸质馆藏资源为中心、以业务管理为中心、以服务用户为中心的转移路径^[13-14],根据功能完善程度的不同,书目的功能实现分为三个层次^[15],由低至高依次为固定化的人工被动式服务、自主学习式的查询式服务、主动积极的精准个性化服务

早期书目工作服务以传统的纸质书籍目录为代表,

预先制定固定的书目工具,用户或工作人员再根据具体需求按图索骥获得文献,重在发挥书目的索引服务功能,属于初级层次的人工被动式服务。

进入20世纪90年代中后期,书目资源的主体从传统馆藏向网络虚拟资源延伸和拓展,极大拓宽了书目服务的空间,数字书目由此而生,其功能由传递具体书籍逐渐转移到满足用户的知识需求,以挖掘组织信息资源为主要内容,由文献传递服务转向知识导航服务^[16],属于第二层次的启发式服务方式。

总结而论,传统书目、数字书目多从信息资源角度出发(见表1),重视对书籍的组织加工、高效检索、固定组织以及范式分析,呈现结果多为原始信息的筛选重组、二次信息的标准化生成,这虽然在一定程度降低了用户获取所需知识的付出成本,但忽略了用户的知识背景与深层次知识需求,未考虑用户使用书目工具的能力和对呈现信息的接受效果。

表1 不同类型书目的区别

书目类别	研究对象	研究目标
纸质书目 ^[17-18]	纸质书籍资源	序化组织资源,创建索引信息
数字书目 ^[17,19-20]	数字书籍资源	实现资源在语义层次的知识提示与组织
智能书目	数字书籍资源 用户信息	通过最符合用户能力的途径满足用户最真实全面的资源需求

网络环境下书目面向的用户更为广泛,其提供的功能急需扩展与延伸,这是信息时代发展的需要与必然^[21]。随着原始文献信息的爆炸式增长,不以用户需求及知识背景为导向的书目工具,常需要用户扩充自身知识背景,付出较大的时间精力成本才可将书目提供信息内化为知识。于是,图书馆管理系统从以业务为中心的第二代系统向以服务为中心的第三代系统更迭^[22],出现了诸如标引智能书目、智能书目推荐、智



能检索、智能阅读等一系列提高书目工具使用效率、考虑用户需求和知识背景的书目工具，但其大多仅就标引、推荐、阅读等单方面以用户为导向，或以实现高效的自动化服务为目的，而未系统化地将“以用户为中心”的服务理念作为贯穿书目编撰的核心内容。

本文在此基础上，提出了“智能书目”的概念，推动书目服务迈向第三层次，将书目工作中心导向更多地转变为用户，从用户知识需求角度和知识背景出发，力求服务对象的普及化与服务内容的个性化，可有效缓解日益繁杂的书目资源库、渐趋多样的用户需求和有限的用户知识背景三者之间的矛盾，实现三者交集的最大化。

1.2 智能书目概念

“智能书目”一词在以往的书目工作中很少被单独提及，常见于以“标引智能书目”“智能书目推荐”等组合词形式出现，其主要是通过将人工智能技术引入书目工作中，实现书目工作的自动化，或就书目某一项功能展开智能化的实践探究，未针对性地对“智能书目”界定概念。在此基础上，本研究提出了“智能书目”的概念，认为智能书目是指依托人工智能和大数据分析技术，根据不同类型用户的知识背景和需求，对文献进行标引、组织、关联，为不同用户提供相宜知识服务的信息集成平台。

基于用户视角、根据用户学习能力组织资源信息^[23]有利于改善用户使用书目的体验感，“以用户为中心”作为智能书目的核心思想，蕴含在书目构建的每一步骤中，体现在书目的标引应纳入用户需求而多元分类，书目的组织应基于用户目的而动态组编，书目的检索应考虑用户能力而人性服务，书目的推荐应依据用户境况而精准推送，书目的阅读应按照用户视角而循序渐进，进而构建以服务用户为目的的书目。

1.3 智能书目构成

书目的构建在于服务用户，了解用户有利于书目构建的完备性，通过收集用户需求、用户知识背景、用户学习能力、用户知识获取障碍、用户检索习惯、

用户兴趣爱好等用户信息^[24]，为书目的服务方向、服务程度导航。鉴于用户知识背景一定程度上体现着用户学习能力、用户知识获取障碍等信息，而用户兴趣爱好、用户检索记录作为用户知识背景外延与用户需求的交集，故可将书目对用户的研究聚焦于用户需求和用户知识背景，有利于精简用户信息，提高智能书目的凝练度和针对性。

传统书目主要以书目资源库为研究对象，智能书目将研究对象扩展为书目资源库、用户需求和用户知识背景，根据三者之间的交集关系可将书目资源分为四类（见图1），包括用户不需要的一般书目、用户需要但不具备相关知识背景的书目、用户需要且具备相关知识背景的书目以及用户需要的书目资源库之外的书目。用户使用纸质书目、数字书目工具得到的主要为部分不需要的一般书目、需要但不具备相关知识背景的书目、需要且具备相关知识背景的书目，这需要用户进一步的筛选或背景知识学习方能将其内化为知识，但忽视了用户的交互体验，且不能确保满足用户的特定书目需求。

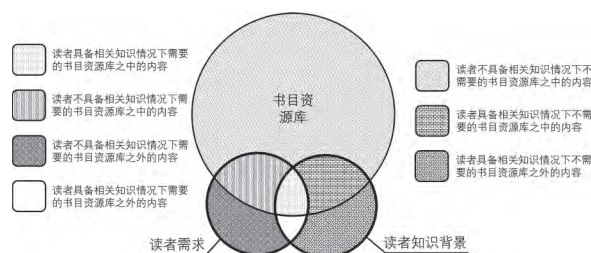


图1 书目资源库、用户需求及用户知识背景三者的交集关系

基于此，智能书目拟通过书目资源库的扩展和组织、用户需求的精准划分、用户能力定位三个步骤，将用户需求进一步融入用户知识背景与书目资源库的交集之中，使三者实现最大化的重叠，保证提供给不同用户的是易于内化为自身知识且属于用户需要的书目资源。为此，智能书目的实现主要包括书目资源库建设、用户需求分类和用户能力定位三个部分。

2 书目资源库建设的组织与更新

2.1 书目资源库组织

传统书目在于著录文献内外部特征,按照一定格式编排组织,揭示和报导文献信息^[18],本质上即是通过各种技术手段筛选、压缩、整序、揭示分散的原始文献信息,形成便于传递的二次描述信息。

当代的书目实践工作相比以往有所改变,一方面是研究对象发生了泛化,由规范的纸质文献信息扩展至非规范的虚拟数字化信息;另一方面,书目工作实践领域已由传统学术科研领域席卷至大众化知识服务环境,面向的用户更为多元化,不同用户知识背景参差、需求千差万别。为调和不同用户的个性化需求,智能书目将不同用户的需求、知识背景等主观性因素融入信息收集、信息揭示、信息组织和书目编撰过程。

智能书目的书目资源库建设主要在精准性、系统性和动态性三个方面体现了智能化的特点:一是将文献的语用信息纳入文献揭示的工作中,可据此形成不同用途的书目提供给相应的用户,助力个性化、精准化的服务;二是以更为系统化的方式组织文献,由此归纳分析得出的知识内容凝练精简,可实现智能导览和新闻播报的快速、高效化服务;三是以一种动态化的方式编纂书目,由用户根据自己的需求编排书目,以一种灵活化的方式提高用户的体验感和满足感。

为赋予书目智能化的特点,首先,智能书目在传统书目工作基础上,将用户需求纳入信息揭示、信息组织、书目编纂的考虑工作之中。例如,通过挖掘文献的内外部特征获得语用信息,捕捉文献中出现的“导读”“入门”“精通”等带有明显用途导向的词汇,分析提取文献的引用信息中蕴含的语用内容等,将某一

领域的文献资源按照用户用途分为引读书目、入门书目、进阶书目、精通书目等,满足不同用户的个性化需求。

其次,智能书目组织文献资源时注重系统性,针对某一特定领域的学科脉络演化进行分析,在归纳总结基础上提供更为凝练、快捷、高效的知识服务。具体过程如图2,选定某一特定领域,按照时间脉络形成包含知识演化进程、学术传承关系、当下研究热点以及活跃学者等内容的知识图谱。通过分析归纳式服务,快速引导用户浏览某一领域的知识演化信息及最新进展,构建智能书目的智能导览和新闻播报功能。

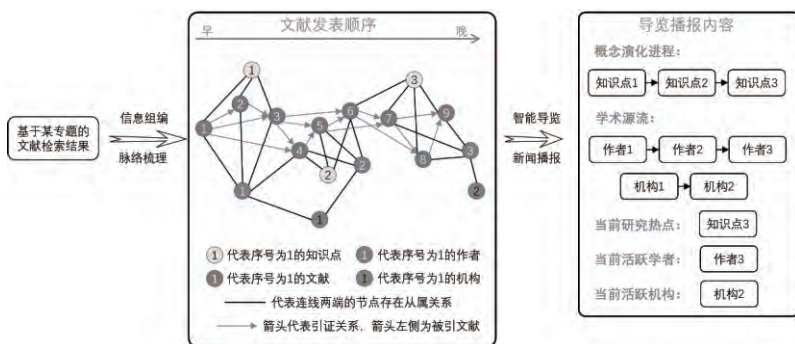


图2 智能导览与新闻播报示意图

最后,智能书目在书目编纂过程中可提供宽松的变动空间,首先是书目组织对象可以多元化(见图3),包括文献、知识点、作者、机构等粒度大小不一、角

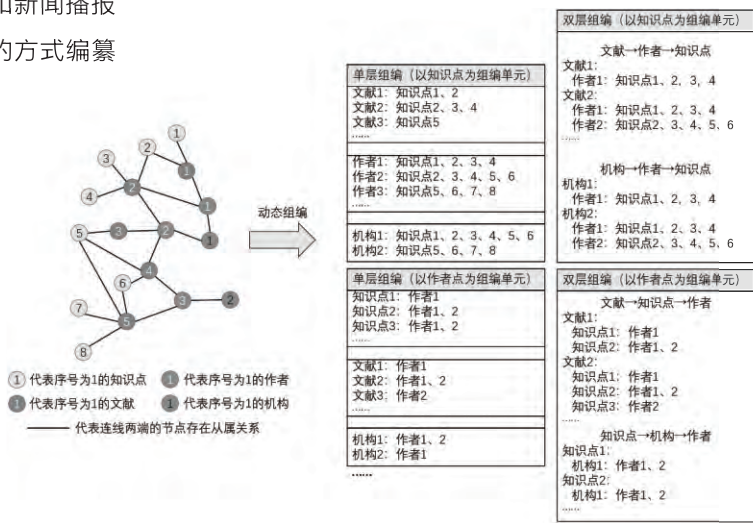


图3 智能书目动态组编示意图



度多样的组编单元。另外,书目组编的层级可指定,可指定2个层级,如指定作者、文献为书目编纂的2个层级,则需罗列每个作者发表的文献;亦可指定3个层级,罗列隶属每个机构的每个作者的文献。

此外,智能书目通过动态组编功能增大书目资源库与用户需求的交集,提供用户更为广泛的选择空间。层级的关系可调整更改(见图3),用户如需了解不同机构在某个知识点的学术贡献,可指定一级目录为知识点,二级目录为机构,找寻对应知识点下辖的机构分布状况;而如需获得不同机构的研究重点,则可将一级目录更改为机构,二级目录更改为知识点,通过更改编目依据的顺序即可实现。

2.2 书目资源库更新

文献资源、用户需求、用户能力呈现动态变化,这使得书目资源库的建设并不能一蹴而就,而是一个动态发展的过程。首先,书目资源库中的文献资源应根据用户直观需求(用户检索记录)、潜在需求(用户浏览记录),通过扩充范围方式增大书目资源库与用户需求的交集。另外,文献资源扩充之后需对相关的书目资源信息进一步揭示、组织和编目,按照智能书目的书目资源库建设方法,及时动态更新书目结构,确保书目资源库的前瞻性和时效性。

书目资源库的建设作为智能书目的基础环节,是提供服务的关键所在。为满足不同用户的个性化需求,智能书目需要一个完备、智能、高效、精简、动态更新的书目资源库。

3 智能书目系统用户分析

2012年,Marshall提出的“图书馆服务平台”^[25]迅速赢得了业界的共识与认可。“服务因你而变”“互联网+”等用户驱动型变革的大时代背景,催生出大量利用大数据和人工智能技术等实现的“智慧服务”形式^[13]。为此,智慧图书馆服务的升级和转型,应以“书”为核心的管理体系,转向以“人”为核心的服务体系^[26]。书目系统作为图书馆系统构建的中心环

节,亦应贯彻“以人为本”的服务理念,将用户分析纳入智能书目系统建设的核心工作。

用户画像构建是分析用户的常用手段,通过以用户为中心,合理划分用户维度,提取用户属性标签,利用用户真实数据结构化地表现用户属性特征^[27]。针对用户画像的构建方法,主要分为两类,一类在于从不同维度用户属性中定量获取高价值信息,进而生成为用户标签,构建用户画像^[28],可直观反映用户静态特征;另一类注重通过提供用户历史数据来模拟用户特征,建立用户概念模型,定性分析用户需求,抽象用户画像^[29-31],可识别用户潜在变化信息。

书目系统用户的分析可结合两种构建用户画像的方法,剖析分类用户需求,把握用户动态发展特点,评估定位用户能力,识别用户静态阶段特征。据此,将用户分析过程分为需求分类和能力定位两部分,可有效兼顾用户静态固化特征和动态发展特征,逼近“用户真容”。

3.1 用户需求分类

智能书目从用户的个性化需求出发,通过不同的服务方式满足用户多样化的需求,通过剖析用户需求的类别做好针对性的服务。从用户角度出发,可将用户对文献的需求大致分为研究型、学习型、应用型、享受型四类^[32],用户使用书目目的分别为从事科学或战略情报研究、学习科学文化知识、职业学习进修及业余创作研究、浏览欣赏。

本文从书目系统建设、服务角度出发,根据用户为满足自身需求而采取对资源使用手段的不同,将用户需求分为检索需求、阅读需求、问答需求以及其他需求。

3.1.1 检索需求

检索需求指用户为解决理解、表达等语言活动中碰到的问题而临时产生的对个别义项的查询需求。智能书目对具体的检索信息进行语义和语用层次的挖掘,为怀有不同检索目的的用户量身定制个性化的精准服



务。针对用户清晰明确的显性需求,智能书目根据用户提供的特征信息全面且精准地搜寻最佳匹配结果;针对用户模糊或潜在的隐性需求,智能书目以智能检索的方式帮助用户构建清晰明确、符合自身需要的规范检索式。

智能书目提供的智能检索功能,具体流程可分为四步(见图4):首先根据用户的活动记录判定用户潜在可能的需求,据此匹配用户的输入检索信息,结合书目资源库与之匹配关联,通过关联匹配结果指引用户输入信息,包括自动补全、校正检测参考以及可能与之关联的词汇等,同时可通过语义挖掘推测用户检索的目的或意图;其次,可依据前一步产生结果构建清晰明确的规范检索式,借此分析用户输入信息的语义内容,更为准确地确定用户的真实文献需求,将隐形需求显性化,促进检索关联匹配层次上升至语义和语用层次;再次,可依据检索结果关联推荐用户可能需要进一步检索的内容,并进一步监测用户的检索行为;最后,根据用户对检索结果的反馈可判断性地产生相应的个性化后续服务,如用户手动重构检索信息或接纳书目推荐的检索信息,则在更新用户活动记录的基础上,开始新一轮的智能检索服务。

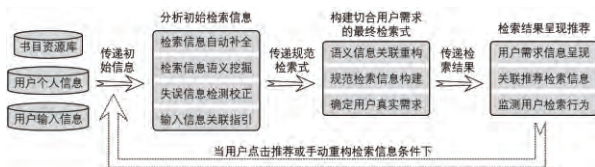


图4 智能检索具体实现流程

3.1.2 阅读需求

传统书目以推荐服务为主,对用户阅读过程中的需求关注较少,但一般用户在阅读之前对于文献的阅读方法、阅读顺序往往欠缺系统性的认识,需要进一步扩充自身知识背景方能无障碍地完成阅读活动。智能书目为缩短用户扩充自身知识背景而消耗的时间,以智能导读的方式为用户指示阅读门径、规划阅读阶段和传递阅读方法^[33]。

智能导读通过搜集用户的个人信息、检索记录和

借阅记录,借助大数据分析挖掘技术,依据用户的年龄、性别、专业、教育程度、阅读爱好、阅读时间、检索记录、下载记录、浏览记录、反馈记录、评论记录,提取用户特征划定不同类别的用户群,生成包含用户的个体特征、阅读目的、阅读行为、阅读习惯等属性^[34]的画像。基于不同用户群的画像,构建用户的阅读兴趣发现模型、阅读行为预测模型、阅读目的挖掘模型等,计算分析得到用户个体标签,为实现精准化的导读服务提供依据。

智能书目意在通过智能导读的形式满足用户多样化的阅读需求,具体流程如图5所示,首先精准化推荐用户需求文献诱发用户阅读兴趣,进而以辅读的提示帮助服务促进用户无障碍阅读,最后通过关联推荐帮助用户进一步理解文献内容,提升阅读层次。其中,引读服务阶段实现资源的个性化、差异化订制阅读,重在引导用户从零阅读到开始阅读,从单一阅读到多元阅读,从浅阅读到深阅读,从浏览式阅读到精细化阅读。辅读服务阶段结合用户知识背景判断用户阅读目的及难点,助力用户在完成阅读的基础上加深对该领域的了解。扩展服务阶段,智能书目可通过关联推荐更深入了解文献或该领域的其他文献进而提供给用户,诱发下一步的阅读活动,实现个性化的指引,帮助用户达到阅读的预期效果。

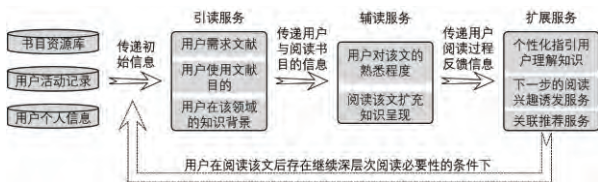


图5 智能导读具体实现流程

3.1.3 问答需求

用户因持有某一确定问题而产生的,在书目中寻求确定答案的需求即为问答需求。传统书目检索得到的大多为一些相关内容的文献罗列结果,不能满足用户的问答需求。智能书目借鉴自动问答形式技术产品^[35],将自动问答系统引入到智能检索的服务之中,用于满足用户的问答需求。



自动问答系统接收自然语言问句，随后返回给用户一个准确的答案。智能书目的问答系统分为问题自动生成子系统和问答检索子系统。问题自动生成子系统负责自动生成问题答案对语料库，这主要借助于建设的书目资源库生成不同专题的问题与对应答案语料库，通过 BERT 模型剖析语料库语义结构生成知识库，进而自动生成问题答案对（见图 6）。问答检索子系统负责语义解析用户输入问题，检索匹配问题自动生成子系统的问题答案对语料库，返回高度相关的最佳对应答案^[36]。

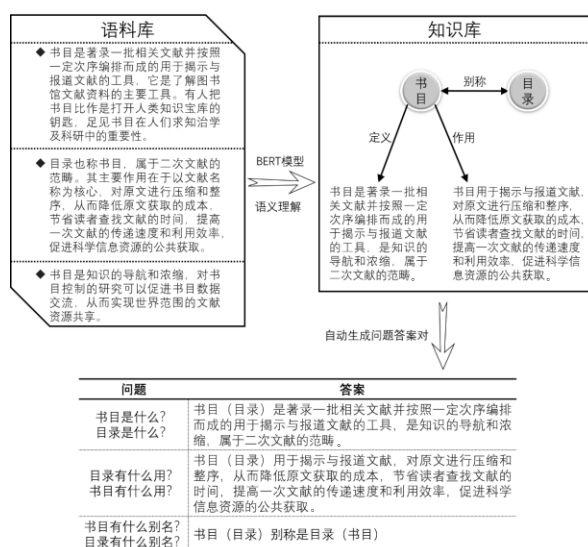


图 6 问题答案对自动生成示意图

通过 BERT 模型和文本标注技术，建立基于深度学习的生成式问答模型和检索式问答模型，提高书目问答功能的效用。智能书目的自动问答系统可避免用户在输入确定问题后进行二次查找或逐一浏览检索结果，以直接给出对应答案的快捷方式，大大节约用户查询时间，达到自动化、准确化、高效化的目的。

3.1.4 其他需求

除检索、阅读、问答三大需求之外，还存在探索式的导航浏览需求、精准化的订阅推送需求、参与式的协同管理需求等特殊性的需求，智能书目在满足主体需求的同时，需进一步将其考虑在内，实现用户各种需求的定制化服务。例如，导航浏览的需求，可通

过用户的检索频次及趋势，将最热门领域或主题的最新进展以新闻播报的形式呈现给用户；订阅推送则可借由依据用户个人信息、历史活动记录的智能推荐功能实现；协同管理则更多以用户的检索需求和评论信息作为书目资源库纳新、重组织的参考依据，实现用户参与书目资源建设管理的目的。

3.2 用户能力定位

为提供恰如其分的精准服务，智能书目将用户的能力因素纳入其考虑范围，根据用户的检索、阅览等历史记录和个人信息定位用户能力，为精准恰当的书目服务提供主观性的参考依据。

3.2.1 检索能力

用户的检索能力作为智能检索的重要参考依据，影响着系统在用户检索过程中提供的帮助服务质量。智能书目利用用户的检索历史记录，可动态评测用户的检索能力变化。利用用户的检索方式^[37]、检索策略、检索词频次分布、阅览等历史记录，可在一定程度判定用户目前的检索能力。

检索方式分为一般检索和高级检索^[38]，用户使用一般检索和高级检索的频次比，可侧面反映用户使用检索工具的部分能力。此外，用户检索策略的多样性表征用户的检索思维能力，检索依据不同可构成多样的检索策略，如关键词、标题、作者、单位、全文、摘要等均可成为检索依据，愈发多样的检索策略，说明用户具备较为周全的检索思维发散能力。再者，根据用户针对相近检索词的检索频次和检索之后的反馈活动，亦可间接反映用户的检索能力。

智能书目通过测定用户检索能力的高低，可用于调整智能检索对用户的辅助程度，尤其是用户检索输入信息的错误识别与校正。如果用户检索能力较差，则更多以智能书目重构的检索式为准，对用户输入信息产生语义关联挖掘，将符合真实需求的备选规范检索式推荐给用户；如若用户具备优秀的检索能力，则更多地考虑用户的具体输入信息，在保证用户输入信息的原始语义条件下，辅之以浅层次的校正，保证检



索结果的精准性。从用户检索能力出发的智能检索服务,在满足用户检索需求的同时,又注重用户的使用体验。

3.2.2 知识背景

现代信息环境下,书目面对的用户千差万别,阅历不同的用户对书目提供的固定化服务的接受理解能力亦参差不齐,实现书目的智能化,则必须了解不同用户的知识背景,提供对应的个性化、精准化的推荐和帮助服务。对用户知识背景的了解,需建立在对知识的分类基础上,据此分析用户日志、用户注册信息、用户浏览历史等,以获取用户在不同领域的认知程度分布状况。

以图7为例,根据一级学科、二级学科以及下辖研究方向可将知识分门别类,继而根据用户对不同领域的认知程度大小可分为精通、掌握、进阶、入门、了解和陌生六个层次,通过分析A、B、C三名用户的浏览信息、检索信息以及个人信息,挖掘不同用户检索和阅读文献的语用信息,判定A、B、C用户在

一级学科、工学下辖二级学科和计算机科学研究方向这三个不同层次的知识背景分布状况。此外,亦可根据其他知识分类体系划分知识类别和层次,定位用户对不同层次知识类别的掌控程度,从综合、全面、系统的角度定位用户的知识背景水平,确保定位结果的准确性。

定位用户对不同知识领域的熟悉程度,可用于判断用户的需求变化及趋势,进而有助于精准化的关联推荐,诱发用户阅读兴趣,从而因材施教实现智能导读。此外,用户知识背景的定位亦可作为智能检索的参考依据,从用户知识背景动态发展阶段出发,有利于透析用户真实需求,构建最符合用户需求的检索式。

4 智能书目系统的服务分析与开发实现

智能书目的最终呈现形式表现为向不同用户提供相宜知识服务的信息集成平台,“以用户为中心”作为智能书目的应有之义和核心内涵,应体现并融入在基于用户数据驱动的智能书目系统平台搭建过程中。

4.1 用户能力与需求的交叉关系分析

基于用户数据驱动的智能书目系统构建注重实现动态组编、智能导览、新闻播报等多样化功能,拓展用户使用书目系统的选择空间,增大书目资源库与用户需求之间的重合度。首先需要通过信息采集手段,探明用户检

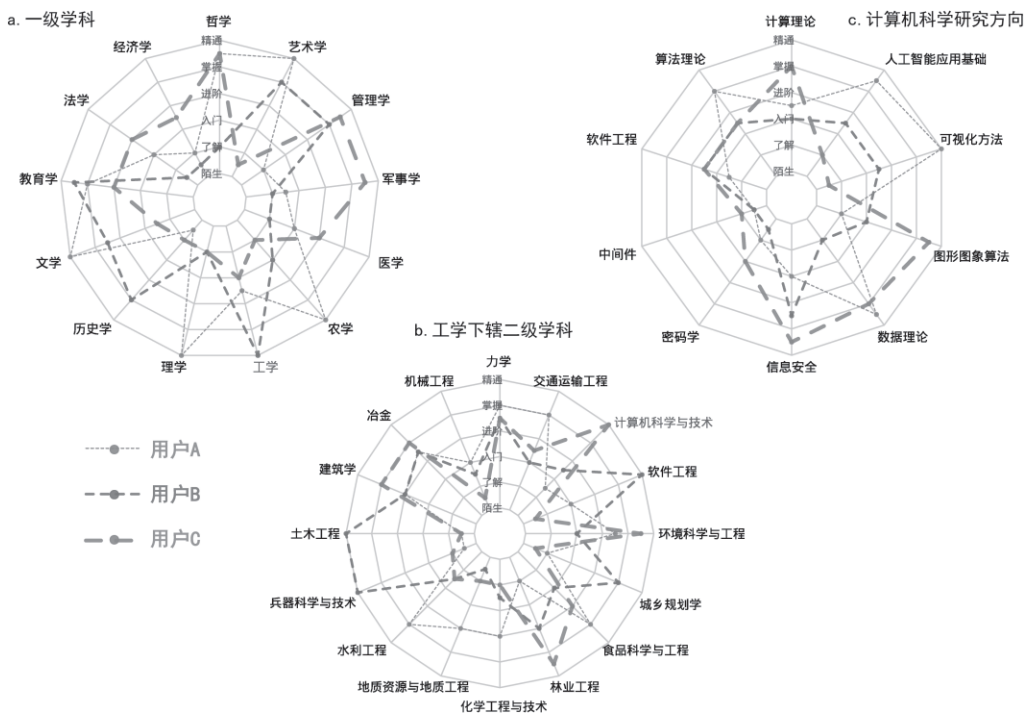


图7 用户知识背景评测



索能力和知识背景, 确定用户能力范围大小, 继而结合用户的各种需求, 确定二者之间的交叉关系; 之后, 智能书目系统的服务方向、服务策略、服务程度、服务方式均依据二者之间的交叉关系作出对应的决策。

当用户能力与用户需求在某一领域存在交集时, 说明用户在该方面的需求有一定知识背景, 智能书目系统在智能检索方面将用户输入信息设置较高的权重, 保证用户输入信息的完整性, 同时更多地推荐该领域新闻播报内容, 便于用户及时掌握前沿进展; 其次, 在导读方面将深层次的文献推荐给用户, 在导读过程中适当减小该领域的基础知识提示服务力度, 更多专注于前沿知识的报导和提示, 改善用户阅读体验。

当用户能力与用户需求在某一领域不存在交集时, 说明用户在该领域尚属新手, 可能由于兴趣或学习而对该领域产生探知需求。为此, 智能书目系统在智能检索方面适当减轻用户输入信息的权重, 在用户输入信息的语义基础上构建规范的检索式, 同时更多地推荐该领域的经典文献和入门书籍, 便于用户掌握该领域的基础知识; 其次, 在导读方面注重由浅入深, 先易后难, 在智能导读过程中, 增强书目在该领域的知识提示服务力度, 帮助用户温故而知新, 同时扩大用户知识面, 关注用户在无障碍阅读过程中的获得感。

4.2 开发流程

智能书目拓展传统书目工作的文献记录、文献揭示、文献组织、书目编纂、书目控制、书目情报服务六步过程,

将精准满足用户需求和定位用户知识背景加入书目建设工作之内, 由智能化技术手段实现, 在系统构建过程中展现为信息收集、信息处理与分析、关联评价模型构建、信息存储、信息交互、信息呈现六个模块 (见图 8)。

其中前四者用于构建书目资源库、评测用户能力, 信息交互模块用于获取用户需求并提供相应的指引性服务, 确定书目资源库、用户需求、用户能力三者之间的分布状况及交叉关系, 指明信息呈现模块的服务方向和服务程度, 提供给不同用户个性化的服务, 促进书目资源库、用户需求、用户能力三者之间交集最大化。本文构建的智能书目系统, 操作环境为 Linux, 信息存储数据库为 Mysql、Neo4j, 选用的 Web 框架为 Django。

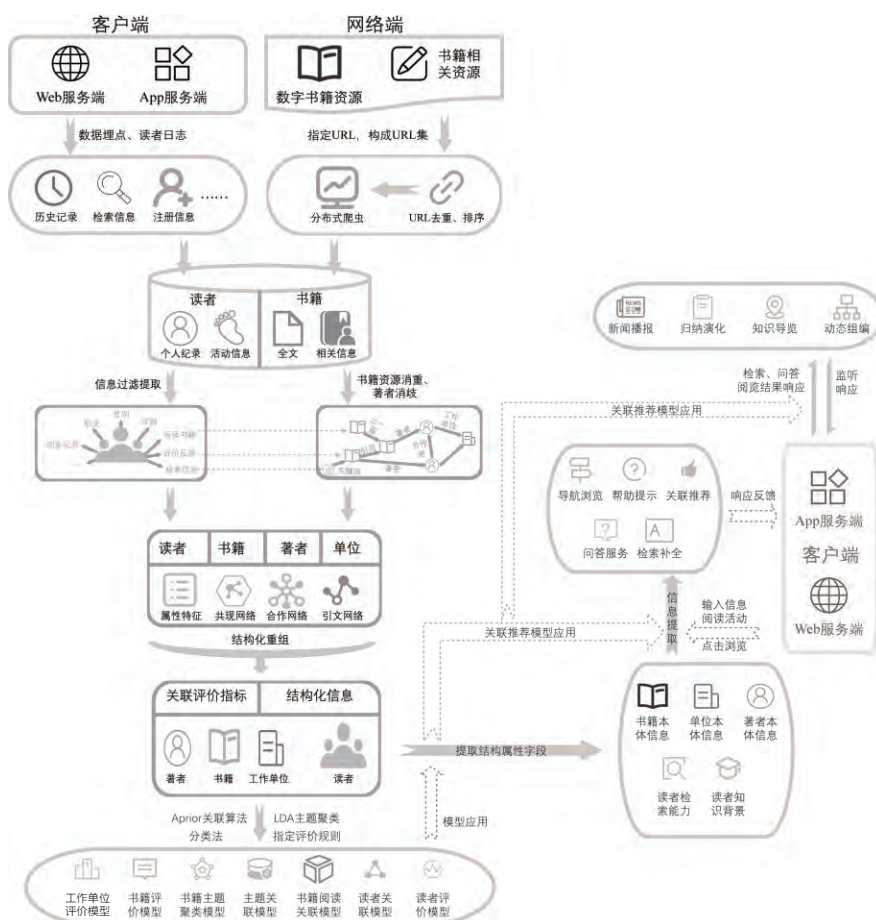


图 8 智能书目系统开发流程



4.2.1 信息收集

智能书目的信息收集模块,包括书目资源库收集和用户信息收集两部分内容。书目资源库的收集通过指定网络端的数字书籍信息源、书籍相关信息源,获取URL集并去重排序,将处理后的URL集传递至分布式爬虫框架,实时采集书籍的评论、简介等相关信息;用户信息的收集主要依托客户端的用户日志和数据埋点技术,获取用户个人记录、用户活动信息等,进而用于评测用户的知识背景和检索能力。

信息收集模块可确保书目资源库建设和用户信息获得的全面性,为书目系统的综合个性化服务提供数据支持,此步为书目资源库建设的物质基础,对书目数据、用户能力的范围限定具有决定作用。

4.2.2 信息处理与分析

信息处理与分析模块分为针对文献信息、用户信息的分析处理两部分。用户信息的分析处理首先通过信息过滤提取高价值的用户个人信息、用户活动信息,获取单个用户的属性结构信息、检阅书籍的关联信息以及多个用户之间检阅书籍的共现网络信息,据此定位用户的检索能力和知识背景,构建用户画像,挖掘用户兴趣,为用户群的关联聚类提供依据,继而为智能书目系统在用户使用过程中提供的帮助具有指引作用,同时在一定程度上可补充文献的价值评价、推荐关联等信息,并可结合文献的引用信息构建评价指标,进而形成多元化的书目文献评价体系。

文献信息的处理分析,通过消重、去歧步骤获取著者、书籍、工作单位、关键词等之间的网络结构信息,进而提取书籍之间的引文网络信息、著者之间的合作网络信息、工作单位之间的合作网络信息、关键词之间的共现网络信息,同时提取书籍、著者、工作单位的属性结构特征,提供书目分类组织的参考依据。该步骤重在规范著录文献的内外部特征信息实现数据的高度结构化存储,保证书目资源库的可操作性、系统性、规范性和实时性。

4.2.3 关联评价模型构建

智能书目的关联模型涉及文献、著者、工作单位、

用户等多个层面的内容,通过创建诸如文献参考引用、作者合作研究、用户活动规律中蕴含的关联指标,借助Aprior关联算法、LDA聚类算法、知识分类方法,训练优化文献之间、著者之间以及用户活动之间的关联模型,进而提供推荐、导引、预测服务。浏览、借阅、检索等用户访问信息,以及文献之间的引证记录,作为多样化评价指标基础,由此训练得出的评价模型,用于计算编目次序的权重,筛选优质信息,提高书目服务质量。

关联评价模型的建立作为信息处理分析的智能化基础,是书目资源库灵活组织、智能推荐的基础,对于书目系统的个性化、关联化程度至关重要。

4.2.4 信息存储

书目系统经由信息收集、信息处理与分析、关联评价模型应用等步骤之后,将规范、详尽的文献、著者、机构、用户等多角度、多粒度信息以唯一且关联化的本体形式表现,并将本体关联网络的节点和连线信息存储至书目资源库。至此,静态书目资源库的建设基本完成,而信息收集步骤仍采用实时监控手段收集用户各类需求,为书目资源库的动态维护更新提供指导。

此外,由信息收集、信息处理与分析过程获得的不同用户的检索能力、知识背景信息亦存储在智能书目系统中,用于确定用户能力与书目资源库的交集关系,为不同用户获得个性化服务提供决策参考依据。

4.2.5 信息交互

信息交互用于向书目系统传递用户需求,主要分为用户活动和系统响应两个部分,其中用户活动包括检索、阅读、浏览、问答等,由关联模型兼顾挖掘用户需求 and 定位用户检索能力,匹配与用户相宜的书目资源,实现智能检索、智能导读、智能问答,具体展现为检索信息的提示指引、纠误补全、关联推荐、盲点提示、阅读指引、问答匹配等指导、辅助、答疑服务。除此之外,针对用户输入信息的语义内容匹配关联存储信息,系统响应后输出检索结果,之后提供语义、语用选项供用户进一步筛选或组织。

书目资源库与用户之间的交互程度彰显书目系统



的智能性,即以指引提示解用户之惑,借关联推荐开治学之路,提高改善用户检索资源的效率及体验,增强书目的治学门径功能。

4.2.6 信息呈现

信息呈现作为书目系统的最后一环,重在为用户提供多样化的书目资源组织方式以及关联推荐服务,组织方式的灵活性给予用户广大的选择空间,增大书目资源与用户需求之间的交集。智能书目打破固定罗列呈现方式,针对检索结果的组织形式力求多样,可指定检索结果的组织形式,具体选项包括时间、地点、文献类型等;此外,可指定文献组织单元的内容大小,如学科领域、研究专题、关键词等;亦可指定创作者组织单元的规模大小,将创作者进一步细分为作者、

机构、团队等。

再者,智能书目以动态组编方式满足用户特定需求,通过信息处理分析模块挖掘汇总检索结果,将知识演进历程通过智能导览形式呈现,同时结合用户的个人信息,关联推荐用户潜在的需求内容。依据不同组织对象形成的书目可以相互嵌套,不同层级书目可依照用户要求进行次序更改,在很大程度上赋予了书目十分宽松的变动空间。

4.3 验证呈现

智能书目系统构建样例,以南京大学图书馆约 28 万本纸质馆藏图书书目数据展开,如图 9 所示,现已实现功能主要包含检索、导航、阅览三项,分别用以满足用户不同需求。书目系统的检索模块在收集用户

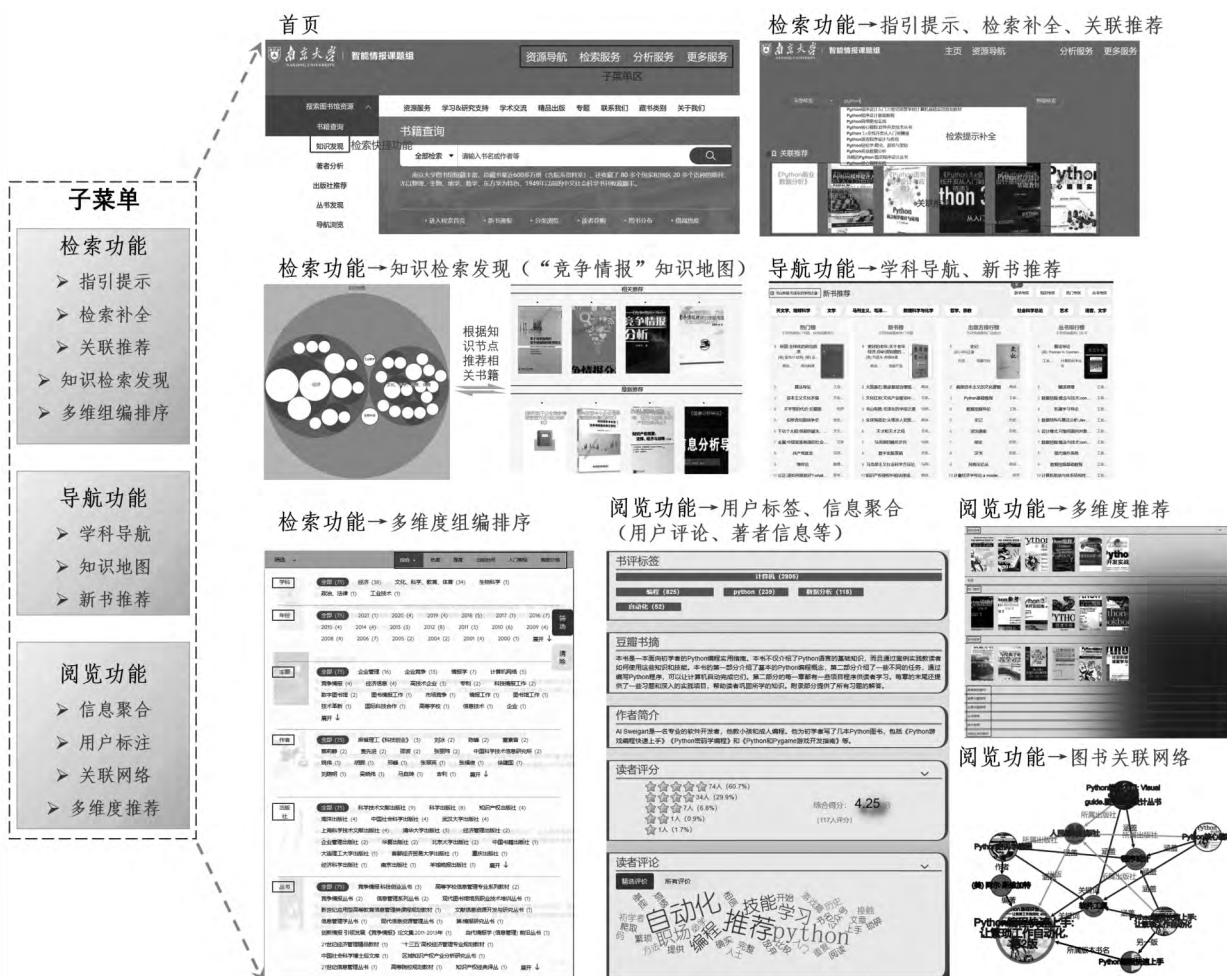


图 9 书目系统各功能模块运行界面



数据基础上,提示、指引用户构建规范检索式,实时推荐高度匹配用户需求的相关图书;在凝聚图书知识属性的主题词信息构建的知识库基础上,实现知识的检索发现,获取某一领域交互式知识地图,根据用户对知识节点的点击浏览行为推荐最新相关图书;根据图书内外部特征生成检索结果的组别,用户可选定拖拽组别选项,实现检索结果的动态组编,并可就具备连续性的图书特征选项(如出版时间、厚度、价格等)展开排序。

书目系统的导航模块,包括新书专区、知识专区、丛书专区、热门专区四个子模块,新书专区集中于呈现不同学科下辖知识点的最新图书,知识专区聚焦于根据某一学科知识地图查询最新相关图书,丛书专区致力于筛选不同学科下辖知识点的热门丛书,热门专区专注于以用户借阅量挑选当下最符合用户需求的高借阅量图书。同时,不同学科导航专区,根据图书摘要、简介等蕴含的语用信息,实现入门书籍、教程书籍推荐等功能。

书目系统的阅览模块,用于呈现检索详情页,根据书目信息采集用户书签、用户评论、著者介绍、丛书简介等相关信息,利用信息聚合功能处理同源异构数据,全面解析图书内容;以图书内外部特征为关联节点,构建完整的图书关联网络,在检索详情页就图书页数、价格、出版社、著者等形式特征实现图书推荐,根据图书主题词、学科属性等内容特征挖掘相关图书。

5 结 语

书目作为图书馆服务的基础,在以人为本、由新兴技术支撑的智慧图书馆建设中意义重大。为此,智能书目贯彻以用户为中心的思想,由用户数据驱动,落脚于优化书目资源库的灵活个性化建设、用户需求的精准针对性分类、用户能力的全面综合性定位三者之间的关系,为书目管理系统的智能化服务提供了一条现实可行的路径,有助于智慧图书馆知识服务平台的个性化、动态化、智能化建设。据此构建的智能书

目系统,主要具备以下优势:

(1) 书目资源库在服务精准性、组织系统性、展示动态性、更新实时性四个方面具备智能化的特点;

(2) 根据用户注册信息提取用户静态特征,结合查检阅览行为捕捉用户动态特征,确定用户知识背景、用户检索能力,从检索、阅读、问答、导航等多个需求角度提供针对性的服务;

(3) 通过指引补全式检索、个性优质化推荐、启发引领式导读、分析归纳式导览、多样动态式组编、新闻播报功能,力求得到用户的个性化需求、有限知识背景和书目资源三者之间的最优解,为每个用户打造属于自己的专属书目。

智能书目的建设是一个长期的、不断完善的过程,其中各个模块的构建仍有拓展、延伸、改良的空间,后续有待于针对专一特定功能展开纵深或定量化的研究,如构建评估用户能力的指标、优选作者姓名消歧的方法、探析用户需求阶段划分的途径等;另外,智能书目应具备开源性的特点,现有的智能书目系统仍具有横向发展的潜力,用户需求的多元化可为智能书目功能的横向扩展提供参考,助力提高智能书目系统建设的全面性。

参考文献:

- [1] 柯平.“后评估时代”公共图书馆的战略重点与发展方向[J].图书馆论坛,2019,39(7):1-12.
- [2] 初景利,高春玲.新时代图书馆与图书馆学的重新认识——兼论图书馆学教育的本原回归[J].图书情报工作,2020,64(1):25-31.
- [3] 曹海霞,侯新宇,杨洋,等.展望“十四五”,促进智慧图书馆大发展——第二届中国高校智慧图书馆(馆长)论坛会议综述[J].新世纪图书馆,2021(10):93-96.
- [4] 王世伟.未来图书馆的新模式——智慧图书馆[J].图书馆建设,2011(12):1-5.
- [5] 许磊.图书馆系统演变及其元数据管理[J].图书馆论坛,2021,41(10):118-126.
- [6] KINNER L,RIGDA C.The integrated library system:from daring to dinosaur?[J].Journal of library administration,2009,49(4):401-417.
- [7] BREEDING M.Next generation library automation: its impact on the serials community[J].The serials



- librarian,2009,56(1-4):55-64.
- [8] WANG Y,DAWES T A.The next generation integrated library system:a promise fulfilled?[J].Information technology and libraries,2012,31(3):76-84.
- [9] 殷红,刘炜.新一代图书馆服务系统:功能评价与愿景展望[J].中国图书馆学报,2013,39(5):26-33.
- [10] 周纲,孙宇.开创性的下一代图书馆服务平台解决方案——FOLIO[J].中国图书馆学报,2020,46(1):79-91.
- [11] 许磊,夏翠娟.第三代图书馆服务平台的元数据管理——以FOLIO的Codex方案为例[J].中国图书馆学报,2020,46(1):99-113.
- [12] 吴冰芝.我国书目情报服务现状及发展趋势[J].现代情报,2005(1):43-45.
- [13] 谢蓉,刘炜,朱雯晶.第三代图书馆服务平台:新需求与新突破[J].中国图书馆学报,2019,45(3):25-37.
- [14] 吴建中.走向第三代图书馆[J].图书馆杂志,2016,35(6):4-9.
- [15] 文思.《目录学》(第五-十二章)[J].图书馆,1987(3):50-53.
- [16] 谢敏明.如何构建图书馆书目搜索引擎[C].福建省图书馆学会2009年学术年会论文集,福建:厦门理工学院图书馆,2009.
- [17] 郭哲敏.语义网环境下书目信息资源的组织——实践方法与发展方向探讨[J].图书馆理论与实践,2015(9):47-52.
- [18] 彭斐章,陈传夫.目录学教程[M].北京:高等教育出版社,2004:1.
- [19] 刘炜,林海青,夏翠娟.数字人文研究的图书馆学方法:书目控制与文献循证[J].大学图书馆学报,2018,36(5):116-123.
- [20] 詹丽华.我国中文电子图书书目数据揭示研究[J].图书馆杂志,2021,40(10):64-71.
- [21] 彭斐章,付先华.20世纪中国目录学研究的回眸与思考[J].图书馆论坛,2004(6):5-10,57.
- [22] 郭利敏,张磊.FOLIO的技术选型与运营模式研究[J].中国图书馆学报,2020,46(1):92-98.
- [23] 罗良道.基于读者视角的网络学术资源组织[J].图书馆情报工作,2004(3):79-81.
- [24] 王福泉,罗忠凤,过仕明.基于读者隐性知识的图书馆知识管理[J].情报科学,2010,28(2):219-222.
- [25] GRANT C.The future of library systems:library services platforms[J].Information standards quarterly,2012,24(4):4-15.
- [26] 杨新涯,袁辉,沈敏.向服务平台转型的下一代图书馆管理系统实践研究[J].图书馆杂志,2015,34(9):23-27.
- [27] 宋雪雁,张梦笛.晋江文学城原创文学网站用户画像研究[J].图书馆情报工作,2020,64(23):63-74.
- [28] 韩梅花,赵景秀.基于“用户画像”的阅读疗法模式研究——以抑郁症为例[J].大学图书馆学报,2017,35(6):105-110.
- [29] 单轸,邵波.国内图书馆领域用户画像研究的冷思考:困境与出路[J].图书馆学研究,2019(5):8-13,30.
- [30] HUANG K H,DENG Y S,CHUANG M C.Static and dynamic user portraits[J].Advances in human-computer interaction,2012(2012):1-16.
- [31] TANG T,YIN Z,ZOU Y.A method for telecom user portrait modeling[C]//2017 5th International Conference on Frontiers of Manufacturing Science and Measuring Technology(FMSMT 2017).Advances in Engineering Research(AER),Amsterdam:Atlantis Press,2017.
- [32] 周维萍,牛振恒.对《图书馆服务五原则》的商榷[J].大学图书馆学报,2008(4):80-82.
- [33] 费巍,李雅.谈目录阅读功能的演变[J].出版发行研究,2018(6):10-13.
- [34] 陈臣,马晓亭.基于小数据的图书馆用户精准画像研究[J].情报资料工作,2018(5):57-61.
- [35] 周永梅.基于本体的自动问答系统[D].镇江:江苏科技大学,2011.
- [36] 李州.自动问答系统若干关键技术设计与实现[D].成都:电子科技大学,2019.
- [37] 周剑.本科生信息检索能力实证分析——兼论《文献检索》课程改革[J].中国图书馆学报,2013,39(2):121-129.
- [38] 凡庆涛,杨萍,杜赟,等.电子信息资源的使用统计分析与评估——以北京市科学技术研究院 CNKI、万方数据库为例[C]//2017年北京科学技术情报学会年会——“科技情报发展助力科技创新中心建设”论坛论文集.北京:北京市科学技术情况研究所,2017.
- [作者简介]
- 彭贤哲 1995年生,南京大学信息管理学院博士研究生,研究方向为智能目录、大数据分析与技术。E-mail: pengxz_tm@163.com。
- 石进 1976年生,南京大学信息管理学院教授,博士生导师,研究方向为情报学、大数据分析与技术、智能目录。E-mail: shijin@nju.edu.cn。
- 李明 1970年生,南京大学信息管理学院副教授,硕士生导师,研究方向为学术评价、科学计量。E-mail: njulm@nju.edu.cn。
- [收稿日期:2022-10-26]