

计算机技术在信息行为领域中的应用

郑万霞

(重庆商务职业学院 重庆 401331)

摘要 在现有的信息技术条件下,基于用户信息行为的数据分析成为提升服务质量的重要途径。文中基于图书馆用户信息行为评估视角,在详细了解图书馆用户读者标签实现、读者信息行为采集后,构建了对应的评估指标,最后通过评估计算指标的权重值,确定了未来图书馆用户信息服务的基本工作方向。根据研究结果可知,信息行为领域中的计算机技术可显著提升图书馆的数据分析水平,为强化用户服务能力奠定基础,具有一定的推广价值。

关键词: 信息行为领域;计算机技术;高校图书馆管;服务

中图分类号 TP399

Research on Computer Technology and Application in the Field of Information Behavior

ZHENG Wanxia

(Chongqing Business Vocational College,Chongqing 401331,China)

Abstract Under the existing information technology conditions,data analytics based on user information behavior has become an important way to improve service quality. Based on the perspective of library user information behavior evaluation,this paper constructs corresponding evaluation indicators after a detailed understanding of the realization of library user tags and the collection of reader information behavior. Finally,by evaluating the weight value of the calculation indicators,the basic work direction of library user information services in the future is determined. According to the research results,computer technology in the field of information behavior can significantly improve the level of data analytics in libraries,lay the foundation for strengthening user service capabilities,and has certain promotion value.

Key words Information behavior field,Computer technology,University library management,Service

0 引言

现阶段,信息技术与图书馆服务工作的融合已成为行业发展的必然趋势,但从用户的浏览行为与阅读过程来看,个性化、多态化的阅读行为越来越常见,而图书馆也因无法识别用户的阅读需求而导致自身服务方向出现偏差。为解决上述问题,出现了基于行为理论的信息行为技术,与传统技术手段相比,信息行为理论更强调用户的行为表现及影响用户行为的因素,可帮助图书馆优化工作方向,符合行业发展需求。

1 图书馆用户的信息行为研究

信息行为指人们为了满足自己的信息需求而进行的一系列社会活动,其目的是解决人与信息之间的关系问题。就图书馆行业而言,在信息行为领域中最常见的计算机技术为大数据技术,其技术表现主要包括以下几点。

(1)信息搜寻行为。利用大数据技术完成提出需求,选择搜寻主题,形成初步探索观点,明确需求,收集信息,完成

搜寻等过程,通过大数据技术来解决用户对信息的不确定性,并逐渐简化用户通过网络等渠道采集信息的过程。

(2)信息检索行为。当用户具备一定的检索技能后,可利用大数据等现代化信息技术完成信息数据交互。用户在产生信息行为的过程中,必须借助大数据工具,如功能成熟的检索网站。同时,网络上充斥着大量的信息,在用户行为领域应确保检索行为的去伪存真、去粗取精的要求。

2 计算机技术信息行为领域中的应用

2.1 读者标签实现

2.1.1 读者标签设计

完成读者分类是应用信息行为技术的基础。图书馆平台可通过大数据技术来预测读者感兴趣的领域,识别读者的潜在需求,这是提升图书馆的服务能力的重要途径。

一类标签主要包括动态行为标签、身份属性标签、兴趣与心理特征表现等内容。通过大数据日志、网络爬虫等技术,可以获取读者的信息行为状态,并利用系统规则记录用户的阅读偏好、阅读模式及日常活跃度等^[1]。

作者简介:郑万霞(1988—),硕士,助教,研究方向为大数据。

二类标签包含了用户的详细身份资料,包括姓名、活跃时间、兴趣领域、阅读目的、阅读期间的心理状态等,需通过大数据技术来整合用户的上述信息行为。

2.1.2 读者标签数据库

为满足信息行为大数据分析要求,本文在读者标签数据库建设中采用了Hive模式+Hbase数据库的方式。其中,Hive是利用SQL语法结构对存储在HDFS中的大数据进行读写管理的一种模式,在批量处理离线数据时具有良好的效果;Hbase是在NoSQL数据库结构基础上发展而来的新型数据库,其在处理批量数据,提供数据查询服务等维度具有显著的作用。

2.2 读者信息行为采集

读者的信息行为主要分为显性数据与隐性数据两种。

(1)显性数据采集方案。利用大数据技术,可根据用户的属性特征、偏好等判别显性数据。例如,可通过大数据来检索读者的兴趣领域、兴趣类型、阅读目的等标签,在读者自行选择结束后,即可将其上传至信息系统中。同时,通过大数据的多次迭代,可以纠正系统中存在偏差的数据^[2]。

(2)隐性数据采集方案。经大数据记录、检索用户的信息行为后,即可根据记录的用户访问情况判断其活跃度。例如,根据用户每周登录图书馆系统的次数、检索频率、阅读行为等数据,即可区分不同的用户群体。

依照上述方法,本次读者信息行为采集内容如表1所列。

表1 用户信息行为采集内容

采集项目	数据类型	采集项目	数据类型	采集项目	数据类型
读者编号	int	图书编码	int	管理者编码	int
读者姓名	Varchar	图书名称	Varchar	操作日期	Datetime
读者性别	Varchar	图书出版社	Varchar	操作类型	Varchar
读者手机号	Varchar	图书字数	int	管理者姓名	Varchar
读者动态行为	Varchar	图书内容特征标签	Varchar	管理者联系方式	Varchar

2.3 系统架构设计方案

基于信息行为的图书馆数据处理系统的结构如表2所列。

表2 系统基本架构

系统架构名称	具体内容
反馈层	用户信息行为的客观反馈、主观反馈
结果层	展示与用户信息行为相关的内容,并将其呈现在Android终端与其他阅读终端端口
分析计算层	构建用户信息行为画像,提取图书文本特征值,匹配画像与文本信息之间的数据
数据层	实现对用户信息行为的画像数据、图书数据、系统单元数据等的管理

2.4 系统处理流程

基于信息行为的图书馆大数据处理流程,可以从读者与信息两个维度实现。其中,读者在登录系统后可以产生信息行为,系统通过大数据即可逐渐完成对用户信息行为的收集与积累。随着数据数量逐渐增多,读者画像逐渐清晰,直至图书馆完成文本特征提取,最终将用户标签上传至数据库中。之后,用户在产生浏览等信息行为时,会积累新的数据。上述数据可以用于丰富读者画像,最后利用大数据更新的方法修正信息推荐方向。

2.5 数据处理方案

2.5.1 数据清洗方案

通过上述流程,采集到用户的信息行为后,由于初次统计的数据无法保证是规范、可用的,数据处理中经常出现重复、异常、不一致的数据,这会导致用户信息行为识别的结果不准确^[3]。因此,在数据分析前,需要对数据进行清洗,解决数据缺失问题。一般情况下,在信息行为数据分析中,不一致的数据通常表现为数据超出了正常、合理的范围,或部分数据相互矛盾。针对这类问题,可通过分析问题数据字段,确定其中超范围或矛盾数据后,再进行数据修正。

针对重复数据造成的冗余问题,可通过识别主要字段的方式直接删除其中的重复部分;对于有较高价值的数,则可做数据合并处理。

2.5.2 数据权重处理

基于图书馆的实际情况,在信息行为分析阶段,可以通过大数据识别不同用户标签与计算机处理结果的相关性。在大数据条件下,图书馆用户信息行为评价结构如表3所示。

表3 用户信息行为影响因素内容

一级指标	二级指标	一级指标	二级指标
	用户身份		人际网络
个人因素	用户知识背景	情景因素	系统使用度
	用户信息素养		信息需求分析
	馆内服务水平		信息可获取途径
环境因素	馆内网络环境	信息因素	信息可靠性
	馆藏资源		信息完整性

基于表3的用户信息行为,由于读者的阅读类型与兴趣是不断变化的,而这一变化也会直接关系到用户的信息行为。因此在数据处理中,应及时更新过期的数据标签,并借助时间衰减系数等,处理数据中的异常情况,如式(1)所示:

$$W(t) = \text{初始权重} \times \exp[-(\text{冷却系数}) \times \text{间隔时间}] \quad (1)$$

其中,W(t)表示时间衰减系数;exp表示数据更新经验系数。

2.5.3 系统训练方案

系统训练方案可以检测系统面对大量用户信息行为数据时的反应能力,即检测系统在大数据处理中是否具有好的稳定性与准确性^[4-5]。经过系统训练后,可提供大数据

(下转第288页)