

基于大数据的用户阅读行为分析*

徐飞¹, 徐绪堪², 吴广印³

(1. 江苏省广电有线信息网络股份有限公司, 南京 210008; 2. 河海大学信息管理系, 常州 213022;
3. 中国科学技术信息研究所, 北京 100038)

摘要: 在大数据环境下, 科技文献等各类数据爆炸式增长, 却难以提供快捷和有效的科技文献服务, 究其原因, 主要在于科技文献资源组织没有考虑科技文献用户阅读行为。分析国内外用户阅读行为, 不仅要分析科技文献用户阅读行为, 更重要的是对科技文献阅读用户行为知识进行组织和分析, 才能为科技文献服务提供支撑。文章以提高科技文献知识服务的效率为目标, 以科技文献用户阅读行为为研究对象, 从知识组织角度分析科技文献用户阅读行为知识, 借助大数据分析技术, 对用户阅读行为知识组织的准备、获取、关联分析以及服务等组织过程, 总体架构科技文献阅读行为知识组织模型, 并把用户阅读行为分析融入到科技文献的知识组织和服务过程中, 将用户阅读行为信息升华为用户知识需求, 最终指导知识提供部门为用户提供有效的科技文献知识服务。

关键词: 大数据; 科技文献用户; 阅读行为; 分析模型; 知识组织

中图分类号: G230.7

DOI: 10.3772/j.issn.1673—2286.2014.12.010

大数据时代, 各类数据扑面而来, 在数据的海洋中人们往往无所适从, 迫切期待科技文献服务部门能够从杂乱繁杂的数据中提取出解决问题的知识, 以服务于大众。作为科技文献服务部门如何驾驭大数据? 如何进一步挖掘数据的巨大价值? 这已成为企业界、学术界共同关心的话题。对于科技文献资源使用和服务也面临同样的问题, 科技文献用户的知识需求不断细化和要求不断提高, 传统科技文献资源组织方式难以满足要求。为了给用户提供针对性的高效知识服务, 重点探讨用户阅读行为知识, 并借助大数据方法和技术对行为知识进行获取、组织和应用, 为主动式科技文献资源服务提供基础和保障。

1 相关文献述评

1.1 大数据利用研究述评

朱建平、章贵军、刘晓藏^[1]在剖析国内外大数据研究和应用现状的基础上, 结合大数据特征和属性, 从统计学角度界定大数据概念, 重新审视大数据时代统计工作思路, 将传统统计质量控制中的事后检验转变为事先预测, 重点是揭示隐藏在其中的有价值的模式和知识。秦晓珠、李晨晖、麦范金、孙卓^[2]分析大数据环境下图书馆资源共享和数据分析的不足, 从数据资源、数据分析技术、数据价值性等角度阐述基于大数据构建知识

* 本研究得到“十二五”国家科技支撑计划项目“专利信息资源挖掘与发现关键技术研究”(编号: 2013BAH21B02)资助。

服务引擎体系,包括知识来源与过滤层、知识存储层、知识表示与标引层、知识处理与挖掘层、知识检索与反馈层等构成,为图书馆用户提供有效的知识服务。王平^[3]分析目前科技文献具有主题多样、动态性强等特点,对主题发现和演化方法进行分析,并以中国知网CNKI的图书馆学与情报学学科2003-2013年发表的论文进行验证。薛文静、孔岩^[4]认为大数据环境下图书馆应该把读者的阅读行为、身份特征、个人爱好与习惯和社会关系等隐私数据作为图书馆数据采集的对象,同时提出措施避免读者个人隐私数据的泄露和侵犯。陈臣^[5]在分析智慧图书馆的内涵基础上,认为大数据技术是从图书馆海量、复杂、实时的数据资源中,通过知识发现、价值挖掘、智能提升和价值创造,以大数据应用和服务为核心,以用户需求为主导,构建基于大数据的图书馆个性化智慧服务体系,实现资源和服务的高度融合,最终为用户提供个性化智慧服务。王天泥^[6]分析大数据时代给图书馆阅读推广面临挑战和机遇,最终读者拥有决定阅读行为的主导权和影响力,将大数据技术“3A5步”法应用于图书馆阅读推广中,充分分析读者个人情况、阅读偏好、反响反馈等详细数据,为优化图书馆阅读推广活动提供依据,提供了智慧的阅读推广方案。

1.2 知识组织研究述评

“知识组织”始见于1929年美国图书馆学家布利斯的专著,并在图书馆学、情报学的分类系统和叙词表研究基础上发展起来,是信息组织的最高形式。Budin^[7]总结KOS的功能主要包括组织和保存大量文献内容的工具、信息系统的组成部分等。Pastor-Sanchez^[8]将SKOS与其他的词典、叙词表表现方式进行比较,并建议从用户视角来管理、检索等操作,而且提出从用户需求角度研究SKOS。曾蕾^[9]认为KOS可为一个领域内语义结构建模,并为标签、定义、关系和性质提供语义、导航和翻译,嵌入到网络中帮助用户发现和检索知识。Mai^[10]从认知视角全面分析了人的信息行为对设计受控词汇的影响,将用户来源和价值观作为一个维度进行分析,研究中结合用户需求来描述和组织知识。

1998年,王知津^[11]在《情报学报》上撰文指出知识组织顺应了信息社会和知识社会的发展,梳理了知识组织的定义,并对知识组织的目标及其活动领域进行了深入探讨。张文亮、徐跃权^[12]从微观、中观和宏观三个层次全面系统地理解和解释知识组织的概念和理论

体系,通过知识发现、知识描述最终实现知识的表达、共享与创新。贺德方^[13]从知识组织体系的构建和应用角度,对知识组织体系之间的转化、映射以及标准化等方面的研究成果进行梳理和总结,认为用户参与和用户使用优先是知识组织的趋势。毕强^[14]分析知识组织系统的相关研究成果并揭示知识组织系统构建从机器可读到机器可理解的发展趋势。贺德方^[15]从知识组织体系的构建和应用角度总结出用户参与和用户使用优先是知识组织的趋势。毕强、滕广青^[16]从语义角度研究数字图书馆中的知识组织目标和内容,探析知识组织体系演化路径不断柔化和复杂化。马费成等^[17]针对网络信息资源集成中的困难,提出基于关联数据的网络信息资源集成框架。

王曰芬等^[18]研究了面向个性化服务的知识组织机制,将用户需求和用户隐性知识纳入其中,提出个性化服务的知识组织过程和方法。以侯汉清^[19]为首的南京农业大学研究团队对知识组织系统、受控词表互操作、叙词表等进行了全方位、多角度的研究。夏立新等^[20]从行为学角度研究政务门户知识组织,对不同类型的信息用户提供相应类型的个性化服务,形成面向用户特征信息服务和知识门户相结合的政务门户系统,对面向用户个性化服务在政务门户知识组织中的实践应用。

1.3 用户阅读行为研究述评

对分析和利用用户行为信息来实现知识服务的研究已经有十多年的历史。Cooley等(1997)将基于数据挖掘的网络知识发现分为网络内容挖掘、网络结构挖掘和网络使用挖掘。网络使用挖掘指分析用户行为信息,这些信息产生于用户和网络交互的过程中,分析者可以从用户与网络交互的海量信息中发现规则和模式。白晨、甘利人^[21]借助Bush-Mostertler和Blrgers-Sarin两个经典强化学习模型剖析科技用户信息检索过程中强化学习机理,通过实验数据分析表明强化学习可以成功反映科技用户对数据库检索方法的学习过程,而且Blrgers-Sarin模型拟合效果更好。刘日升、杨振力^[22]从电子期刊、电子书和电子阅读工具三个方面分析国外用户电子阅读需求和行为的研究现状,针对已有文献数据从研究对象、研究切入点、研究方法和研究结论四个方面分别总结研究的特点,不同的用户群体的研究目标和影响因素不同,但每个研究的结论都不能适用于假设条件之外的其他目标用户群。刘亚、蹇瑞卿^[23]

从阅读需求、阅读寻求、阅读处理与使用三个阶段研究大学生的手机阅读行为,并利用威尔逊信息行为模型分析问卷调查,分析大学生手机阅读信息行为的发展和变化规律,提取大学生阅读行为具有缺少对资源和技术了解和关注、没有主客体区分的阅读和传播、轻松阅读和互动阅读等特征,结合干扰和激励阅读的因素提出合理化的建议。钱鸥、李翔翔^[24]从用户体验的视角出发,对综合性高校图书馆的移动阅读服务进行实证性研究,以武汉大学移动图书馆为例,从阅读时间、获取方式、阅读内容、阅读便利四个维度研究存在对图书馆移动阅读服务了解少、阅读碎片化、对休闲娱乐类资源粘性强、对学术性内容阅读较少等行为特征,并为综合性高校图书馆移动阅读服务设计优化提出合理化建议。赵杰、何洁、王曰芬^[25]在文献研究基础上,将信息用户科技文献下载行为作为研究对象,构建科技文献下载行为的构成、要素之间相互作用的下载行为决策概念模型,通过问卷调查收集南京理工大学信息用户数据,运用PLS方法对概念模型进行实证验证,得到如下结果:下载动机、认知和能力因素对行为主体科技文献下载决策有直接的正向影响,人格情绪、信息需求、外在因素和经验对行为主体科技文献下载决策存在间接的正向影响。高春玲、卢小君^[26]从设施特征、用户阅读行为等方面揭示用户阅读图书馆电子资源的使用意愿,以辽宁师范大学师生为目标群体,提出图书馆应用用户移动阅读行为有效引导的策略。

综上所述,目前国内外学者分析大数据环境下科技文献组织的内涵和特征,为科技文献组织带来新的挑战和要求;已有研究成果主要集中在先组式的知识组织体系和方法研究,虽有考虑用户需求和行为的知识组织体系,但难以摆脱先组式知识组织的局限性;已有研究成果侧重以用户阅读行为为研究对象研究用户阅读行为影响因素、电子资源使用意愿,侧重于被动的信息检索服务,缺乏深入分析用户问题,离主动的知识服务还有很大差距,导致无法满足用户的需求。因此,本文重点探讨科技文献用户阅读行为构成,构建用户阅读行为知识的组织框架,借助大数据技术和方法,将用户阅读行为信息升华为用户知识需求,并引导科技文献进行主动式针对性的组织,从而为用户提供有效的知识服务。

2 用户阅读行为构成

用户阅读行为是指科技文献用户在使用科技文献

系统过程中,为得到解决问题的知识(信息)所采用的信息获取方法、信息查询步骤、选用的检索词以及对检索结果选择等行为。为了更好地为用户提供高效知识服务,在目前大数据环境下,传统服务模式难以满足大数据环境下用户对知识服务的需求,因此为了提高科技文献服务效率,通过分析用户阅读行为,获取用户行为信息并进行组织和处理,为文献索引和知识服务提供决策和指导。

用户阅读行为主要包括用户类型、用户问题、用户知识需求、阅读内容、获取方式、用户评价、信息推荐、文献资源等部分构成。如图1所示大数据环境下用户阅读行为结构图。

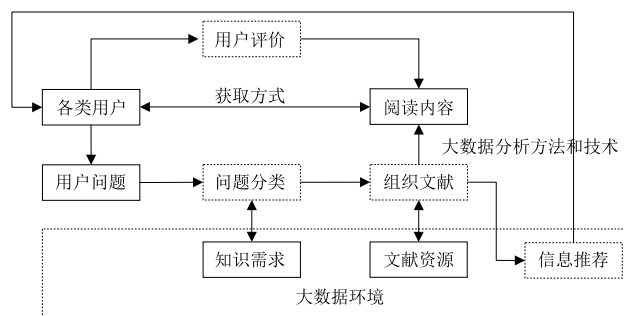


图1 大数据环境下用户阅读行为结构图

在大数据环境下,不仅科技文献资源快速增长,而且知识服务面对的用户是广泛的,不同类型的用户对知识的需求也是不一样的。针对不同类型用户的知识需求收集和分析,形成用户类型库,以用户需求为驱动,按照不同类型有针对性地对科技文献进行知识组织和服务,是科技文献服务系统提升服务质量的重要保证。例如科技文献知识服务面对的用户有不同文化程度的用户、不同知识背景的用户、不同职业的用户、解决不同类型问题的用户等等。结合不同类型的用户的需求,形成对应的知识需求库,借助大数据分析方法和技术,对科技文献资源进行检索和组织,从中提取用户需要的资源形成用户阅读内容,同时结合文献资源利用和检索过程数据采集和分析,提炼出阅读率高的文献资源,并以信息推荐的方式推送给不同类型的用户。在用户阅读文献之后,不同用户对已经阅读的文献资源进行反馈和评价,形成用户评价库,为知识需求提炼和细化提供支撑,同时也为科技文献资源组织提供引导。最终文献资源、用户信息、评价反馈、知识需求等大量数据经过集成、整合之后,借助大数据方法和技术,提炼出用户阅读行为的规律,为科技文献资源组织提供科学依

据, 并能为用户提供有效的知识服务。

3 用户阅读行为知识的组织框架构建

为了解决不同类型用户的不同问题, 除了传统被动的文献资源组织外, 还要通过用户阅读行为知识来提取用户知识需求, 所以需要对用户阅读行为信息和知识进行组织, 针对科技文献资源使用和组织, 通过用户阅读行为知识组织准备、获取、处理以及应用等四部分来完成用户阅读行为知识的组织和应用, 构建一个科技文献用户阅读行为知识组织框架, 针对不同领域的用户知识需求, 主动式组织科技文献资源, 并以推送的方式为用户提供知识服务。用户阅读行为知识的组织框架如图2所示。

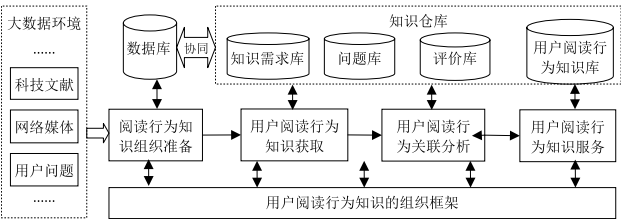


图2 用户阅读行为知识的组织框架

3.1 用户阅读行为知识组织的准备

为了实现用户阅读行为升华为用户知识需求的目标, 用户行为知识组织工作的前期准备是十分重要的, 它是用户阅读行为分析的基础和保证。在用户阅读行为知识组织的前期构建准备中, 需要用户阅读行为概念、类别等基础数据资料等, 虽然国内外学者对用户阅读行为进行大量研究, 但在实际知识组织过程中对用户阅读行为知识组织准备尚未引起足够重视。因此, 欲将从用户行为信息中发现的知识有效地组织起来, 对于科技文献资源的用户, 可以通过纸质和电子资源提供服务, 用户阅读行为各类信息分布在这个服务全过程中, 用户在使用科技文献资源过程中所涉及的各类信息。具体包括: 用户在一段时间内登录网站完成特定检索浏览任务的会话 (Session) 信息; 用户检索和浏览的目标个体——文档 (Document) 信息; 用户注册的账户 (Account) 信息; 用户在检索、浏览资源项目时所使用检索词 (Query) 以及用户在查询和使用这些资源项目时对其有用性的排序 (Rank)。以上信息域是进行用户

阅读行为知识组织的基础, 科技文献资源服务提供设计者应通过客户端数据、服务器日志或数据库等方式记录这些用户阅读行为信息, 作为用户阅读行为分析的基础。每个用户阅读行为按照以下信息项的概念定义。

会话 (session): 是一个客户与服务器之间的不中断的请求响应序列。对客户 的每个请求, 服务器能够识别出请求来自于同一个客户。当一个未知的客户向Web应用程序发送第一个请求时就开始了一个会话。当客户明确结束会话或服务器在一个预定义的时限内不从客户接受任何请求时, 会话就结束了。

文档 (Document): 指网站所提供的网页或信息单元, 是信息的载体, 并可以分类和索引。每个文档应当有一个唯一的URL地址, 为了更好地实现用户行为信息分析, 网站应该首先建立好文档数据库, 确保文档的唯一性, 并维护好文档的属性信息。

账户 (Account): 用于认证用户以便于用户使用系统资源。用户账户信息不仅包括用户标识名, 还包括用户的性别、年龄、教育背景等用户属性信息。

检索词 (Query): 是表达用户信息需求和检索课题内容的基本单元。检索词可能是叙词, 也可能是用户自述的关键词和单元词, 表达了用户本次会话的信息需求。

评价: 用户在查询和使用这些资源项目时对其有用性可能有一定的排序 (Rank), 这可能是显性的评分, 也可能是隐性的关注、下载、保存等行为。

所以, 用户阅读行为信息的概念化形式可以表示为 $I<S,A,D,Q,R>$, 用户行为知识则可以通过分析挖掘这五类信息项中两项或多项之间的关联关系得到。表1列出了部分信息项与用户行为知识的关系。

表1 用户阅读行为知识及应用列表

信息项	用户行为知识模型	应用
D, S	文档关注度	热点推荐、趋势分析
D, R	文档评价	资源推荐
D, S, Q	协同过滤知识	个性化推荐
A, S, Q	用户兴趣度知识	用户兴趣建模、信息推送
.....		

3.2 基于用户阅读行为知识获取

用户阅读行为知识获取主要任务是根据用户问题和检索目标资源的过程进行采集、检测、修正、抽取等

过程初步检测和消除噪音数据,合并同类数据^[27],剔除重复记录数据和不可用的数据资源,形成粗粒度的数据清洗框架,有效提高数据的质量,为用户阅读行为知识组织提供可靠资源支撑。其中日志数据是用户阅读行为知识获取最重要来源之一,因此用户阅读行为知识获取重点就是用户阅读日志数据的获取。

表2 清洗处理后的日志标准格式数据库

LogID	Session	Account	Document	Query	Rank
1	S1	U1	D1	Q1	R1
2	S2	U1	D2	Q2	R2
3	S2	U1	D3	Q2	R1
4	S3	U2	D1	Q1	N/A
5	S3	U2	D4	Q3	R2
6	S3	U2	D5	Q3	R1

用户日志数据的预处理,将多格式、非结构化、含冗余数据的日志整理到日志标准数据库中。经过处理后的日志示例数据如表2所示。每一个信息项可以是一维数值,也可以是多维数值,如Session信息中不仅有session编号,还有IP地址、开始时间、结束时间等。Item信息除了名称外,还可能包括分类号、标引、关键词等。Rank评分数据也可以是多维的,表达用户对相关性、有用性的多方面评价。此外,由于网站功能和应用的差异,有些数据可为空的(N/A)。例如一个分类浏览的网站,可能就没有检索词的信息,搜索引擎日志则大多没有用户评分的信息。

3.3 用户阅读行为知识的关联和分析

用户行为知识的关联和分析主要实现用户阅读行为信息关联和用户阅读行为数据的依据,首先需要应用大数据环境下知识关联组织的方法,将用户阅读行为知识存放到用户行为知识库中,如数据仓库、知识地图等等。数据仓库技术可以将散落在多个数据库中的信息有机地关联起来,形成多维度联系。知识地图则以知识单元(人、文献、概念、知识)为节点,通过相互关联实现知识链接。知识聚类组织法可以将原本分散的信息或知识按主题特征等聚合在一起,使聚合在一起的信息或知识具有某种共同特征或关联。

对于科技文献用户,首先从科技文献共享平台中提

取用户阅读行为,从中获取用户检索和阅读过程中取出会话及其对应的文档和检索词信息项 $I\langle D, S, Q \rangle$,通过对其进行关联规则挖掘,可以获取到具有相关性的文献之间的知识 $K\langle D, D \rangle$,这个知识就可以通过一定的知识组织方式存入用户阅读行为知识库,以用于指导科技文献知识服务,以上的过程中用户阅读行为可以通过以下公式表示: $I\langle D, A, S, Q, R \rangle \rightarrow I\langle D, S \rangle \rightarrow K\langle D, D \rangle$,按照图3所示用户阅读行为模块进行数据初步关联。

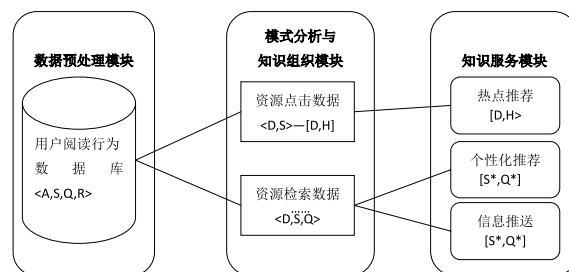


图3 用户阅读行为模块进行数据初步关联

在用户阅读行为知识关联和分析过程中,针对用户问题的类型和特征,根据问题及问题解答相关性来获取所需数据和知识,在用户阅读行为数据初步管理基础上,需要进行深度挖掘提炼成用户的知识需求,同时将用户检索和阅读数据和知识以知识单元为基础进行描述,知识单元可以描述用户阅读行为与不同知识单元之间的关联关系^[28],并针对用户需求与科技文献资源进行关联和映射,经过再组织后形成推荐的文献资源,为用户个性化需求的知识服务提供资源保障,如图4所示用户阅读行为挖掘和关联。

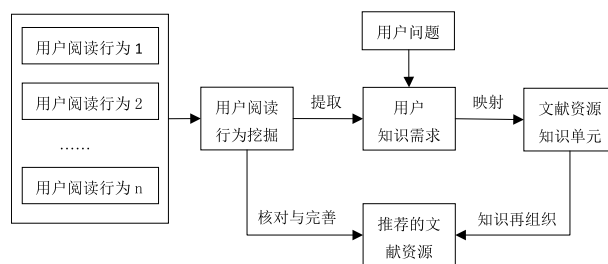


图4 用户阅读行为挖掘和关联

3.4 用户阅读行为知识的服务

用户阅读行为知识的服务不仅仅是通过科技文献资源库来提供知识服务,而且需要结合用户阅读行为,借助大数据环境下可视化方法和技术,为用户提供需要最近研究或者急需的知识服务,减少用户检索的时间。

知识服务的方式可以分为显性和隐性两种方式。显性方式是为用户提供知识的查询和导航,如用户可直接查看当前的热门文档、文档或关键词的访问趋势,可以通过文档的分类快速找到和定位自己需要的文档信息。隐性方式是针对用户阅读行为和知识需求下进行用户所需知识的组织,同时通过提供检索辅助、检索优化、推荐和推送等方式的知识服务,如利用文档索引信息为用户提供检索词补全和更正、检索优化、个性化推荐等知识服务,用户阅读行为知识的服务系统框架如图5所示。语义互操作服务平台是对知识仓库、用户行为模型、知识检索与匹配以及情境感知的知识需求之间进行互操作,是知识服务的核心,通过互操作有助于满足用户的需求,为用户提供问题的最优解,同时结合语义互操作服务平台应用不断修订和完善用户阅读行为库。

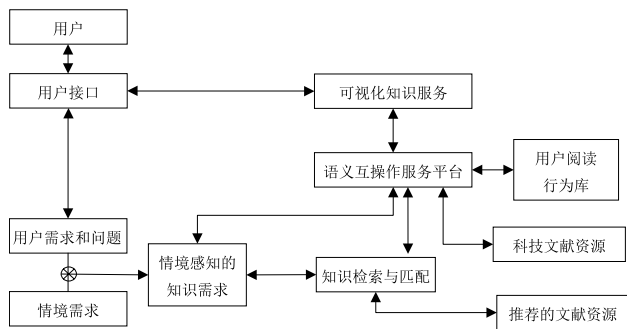


图5 用户行为信息知识组织服务框架

4 基于用户阅读行为的知识应用

用户阅读行为知识的组织和分析目的是为用户提供更有价值的服务,所以通过实际应用反馈来修订和完善用户阅读行为知识组织框架,用户行为是指用户在使用信息服务或知识服务系统时查找信息的过程,一些有经验的、成功的查找案例的过程就是用户的行为知识,将这些行为知识服务于其他用户就是基于用户行为的知识服务。例如,当一位用户到图书馆去借阅一本书时,这本书已被借出,但他认为另一本书完全可以取代这本书,他又提出借阅这本书,并说明为什么借这本书。这时,图书馆员记下了这一借阅过程和替代过程缘由,当另一位用户也来借上一本书的时候,并且也已被借走时,他就可以根据前一用户的借阅行为向这个用户推荐后一本书,从而使这个用户得到满意的服务。

在网络环境下,用户查找信息的行为属于自助行为,一些初级用户或缺乏使用经验的用户非常需要能够

得到一些有助于查找信息的行为知识,这就需要系统能够将成功的查找行为和有效的查找策略记录下来,并在用户需要时提供给他们。当然,用户的行为知识不是简单的检索过程,它需要将众多用户的行为清洗、归纳、总结,整理出有价值的、合理的检索行为过程,并存储于用户行为知识库,最终提供用户服务。用户行为的知识服务不是一个人的行为推荐,而是众多用户行为归纳总结的经验推荐;也不仅仅是用户使用过的检索词推荐,还包括系统通过主题表等相关产生的相关词汇的提供。所以,用户行为的知识服务是用户行为知识与相关知识集成的知识服务。

5 结论与展望

在大数据环境下,科技文献资源增速日益加快,为科技文献知识服务提出新的要求,本文通过探讨用户阅读行为构成,借助大数据分析方法和技术,构建用户阅读行为知识框架,从用户阅读科技文献资源中提取用户知识需求,并依据用户知识需求为科技文献知识主动组织提供支撑,从而为用户提供有效的知识服务。但本文仅仅从总体框架上对用户阅读行为知识进行组织,下一步将对于用户阅读行为知识组织框架实现和应用。

参考文献

- [1] 朱建平,章贵军,刘晓藏.大数据时代下数据分析理念的辨析[J].统计研究,2014,31(2):10-19.
- [2] 孙卓.基于大数据构建图书馆知识服务引擎研究[J].图书馆学研究,2013,18:48-51.
- [3] 王平.基于层次概率主题模型的科技文献主题发现及演化[J].图书情报工作,2014,58(22):70-77.
- [4] 薛文静,孔岩.基于文献分析的国内图书馆大数据应用研究述评[J].农业图书情报学刊,2014,26(11):65-69.
- [5] 陈臣.基于大数据的图书馆个性化智慧服务体系构建[J].情报资料工作,2013,(6):75-79.
- [6] 王天泥.大数据技术在图书馆阅读推广中的应用[J].情报资料工作,2014,(4):96-99.
- [7] BUDIN G. From Terminologies to Ontologies-Advances in Knowledge Organization [D]. Vienna: University of Vienna, 2007.
- [8] PASTOR-SANCHEZ, J-A, MARTINEZ MENDEZ, F J, RODRIGUEZ-MUNOZ, J V. Advantages of Thesaurus

- Representation Using the Simple Knowledge Organization System (SKOS) Compared with Proposed Alternatives [J]. Information Research: An International Electronic Journal, 2009, 14(4).
- [9] ZENG M L. Knowledge Organization Systems [J]. Knowledge Organization, 2008, 35(2-3): 160-168.
- [10] MAI J-E. Actors, Domains, and Constraints in the Design and Construction of Controlled Vocabularies [J]. Knowledge Organization, 2008, 35(1): 16-30.
- [11] 王知津.从情报组织到知识组织[J].情报学报,1998,(3):230-234.
- [12] 张文亮,徐跃权.论知识组织的三个层次[J].图书情报工作, 2011, (1):41-45.
- [13] 贺德方.国内外知识组织体系的研究进展及应对策略[J].情报学报,2010,29(6):963-972.
- [14] 毕强.数字图书馆知识组织系统建构的发展趋势:从机器可读到机器可理解[J].国家图书馆学刊,2010,19(1):12-19.
- [15] 贺德方.国内外知识组织体系的研究进展及应对策略[J].情报学报,2010,29(6):963-972.
- [16] 滕广青,毕强.知识组织体系的演进路径及相关研究的发展趋势探析[J].中国图书馆学报,2010,136(189):49-53.
- [17] 马费成,赵红斌,万燕玲等.基于关联数据的网络信息资源集成[J].情报杂志,2011,30(2):167-169.
- [18] 王曰芬,熊铭辉,吴鹏.面向个性化服务的知识组织机制研究[J].情报理论与实践,2008,(1):7-11.
- [19] 刘华梅,侯汉清.基于受控词表互操作的集成词库构建研究[J].中国图书馆学报,2010,(3):67-72.
- [20] 夏立新,叶飞.行为学角度的政务门户知识组织与整合研究[J].情报学报,2011,28(3):331-336.
- [21] 白晨,甘利人.基于强化学习模型的科技文献数据库用户学习行为研究[J].现代图书情报技术,2009,181(7):85-92.
- [22] 刘日升,杨振力.国外用户电子阅读需求与行为研究综述[J].情报资料工作,2012,(1):97-101.
- [23] 刘亚,塞瑞卿.大学生手机阅读行为的调查分析[J].图书馆论坛, 2013,33(3):97-101.
- [24] 钱鸥,李翔翔.综合性大学在校生移动阅读行为特征研究:基于用户体验的视角[J].图书情报工作,2013,57(22):92-98.
- [25] 赵杰,何洁,王曰芬.科技文献下载行为形成机理及影响因素研究[J].情报杂志,2013,32(4):171-176.
- [26] 高春玲,卢小君.用户阅读图书馆电子资源意愿的影响因素分析:以辽宁师范大学师生移动阅读行为为例[J].图书馆论坛,2014,(2):34-40.
- [27] HERNANDEZ M A, STOLFO S J. Real-world Data is Dirty: Data Cleansing and the Merge/Purge Problem [J].Data Mining and Knowledge Discovery, 1998, 2(1): 9-37.
- [28] 温有奎,焦玉英.基于范畴论的知识单元组织与检索研究[J].情报学报,2010,29(3):387-392.

作者简介

徐飞,男,1983年生,工程师,研究方向:信息分析。

徐绪堪,男,1976年生,河海大学信息管理系副教授,研究方向:知识组织、数据挖掘。

吴广印,男,1965年生,中国科学技术信息研究所研究员,研究方向:知识组织与管理、云计算架构。

Research on Scientific Literature Users' Reading Behavior Based on Big Data Application

XU Fei¹, XU XuKan², WU GuangYin³

(1. Jiangsu broadcasting cable information network Corp., Ltd., Nanjing 210008, China; 2. Dept. of Information Management, Hohai University, Changzhou 213022, China; 3. Institute of Science and Technology Information Research of China, Beijing 100038, China)

Abstract: With the background of big data today, various data concerning scientific and technical literature grow at an explosive speed. However, fast and efficient service is difficult to be provided for the science and technical department, for the reason that scientific literature is mainly organized in a passive way without considering the users' reading habits. On the basis of studies about domestic and international users' reading behavior, not only the scientific literature users' reading behavior is analyzed, but more importantly, the knowledge organization about this behavior is also analyzed so as to provide support to the scientific literature service. Therefore, in order to improve the efficiency of the scientific literature knowledge and with relevant users' reading behavior as subjects, this paper analyzes the reading behavior and knowledge of scientific literature users from the perspective of knowledge organization. By the big data analysis method, this paper makes research on the organization process of users' reading behavior knowledge including preparation, obtaining, correlation analysis and service etc., builds the overall architecture of the organization model, integrates the analysis into the organization and service process, then upgrades the users' reading behavior information to knowledge demands, and finally gives guidance for the knowledge department to provide users with a valid scientific literature knowledge service.

Keywords: Big data; User of scientific literature; Reading behavior; Analysis model; Knowledge organization

(收稿日期: 2014-12-12)