

基于多粒度的图书馆知识服务创新

陈燕方

(中国人民大学信息资源管理学院, 北京 100872)

摘要: 面对检索结果形式单一、语义关联性差的弊端, 以及用户日益多元化和精准化的需求, 图书馆知识服务模式亟待突破和创新。本文在对图书馆知识服务现状梳理的基础上厘清多粒度的概念与内涵, 并提出基于多粒度的图书馆知识服务模型。该模型主要分为四层, 即馆藏资源层、知识组织层、知识聚合层和知识服务层。最后, 详细阐述该模型的三大知识服务功能模块。

关键词: 图书馆; 多粒度; 知识组织; 知识服务; 语义

中图分类号: G250.7

DOI: 10.3772/j.issn.1673-2286.2018.03.005

在科学技术和人文需求的不断发展中, 图书馆已经由传统的物理图书馆, 以及数字图书馆、复合图书馆、移动图书馆等形态逐步朝智慧图书馆的方向发展。按照Web技术发展划分, 图书馆的发展及演变历经图书馆1.0、图书馆2.0, 以及当前的发展目标——图书馆3.0三个主要阶段, 其服务内容与方式也从传统的文献服务、信息服务转向知识服务^[1-2]。

图书馆2.0仍是当前图书馆服务的主流, 泛在的^[3-4]、无障碍的^[4-5]、用户参与的^[5-6]、个性化的^[4,7]信息服务是其主要特点, 但其信息服务存在加工程度低、数量大而质量低、语义关联性差等弊端。尤其在信息爆炸和信息泛滥的大环境下, 图书馆2.0的信息服务面对用户日益精准、个性化的信息需求显得力不从心。针对图书馆2.0信息服务存在的问题, 知识服务及图书馆3.0概念应运而生^[8]。早在2000年, 张晓林^[9]曾提出知识服务是新世纪图书情报的生长点, 是基于分布式多样化动态资源, 贯穿用户实际问题解决过程的个性化、专业化服务。图书馆3.0更是将知识服务作为发展的主要目标, 旨在将无组织的Web内容转变为系统的、有组织的信息, 提供给用户智能的知识服务^[10]。实现该目标的核心任务是对图书馆馆藏资源的组织方式和服务提供方式进行重组。

1 图书馆知识服务现状

国外对知识服务的研究, 起源于以提高企业经济

效益和竞争力为目的的知识管理, 此后该理念被引入图情领域。1997年, 美国专业图书馆协会(SLA)在其会刊*Information Outlook*上专门设立栏目开展对知识管理的研究^[11]; 2001年, Clair^[12](SLA前任会长)明确指出专业图书馆最新发展趋势即开展知识服务, 图书馆员和信息专家应在新形式下为用户提供创新的、获取知识的服务, 使知识服务成为一种信息使用的管理方法。根据知识生命周期划分, 知识服务的流程包含资源的采集、组织、挖掘、创新、分发、利用和反馈^[13]。然而, 在当前网络搜索引擎强大的竞争态势下, 知识服务流程各阶段最亟待解决的问题是资源组织方式。传统的资源组织存在明显的资源异构及语义关联性差的问题, 使用户在检索过程中需要尝试不同的检索入口。此外, 以关键词匹配的模式, 使用户难以从大量的返回结果中获取有规律、有关联的信息或知识。

在当前信息爆炸与信息泛滥的环境下, 用户对网络信息服务的内容、方式及效果的要求愈来愈高, 如以知识为单元的“微信息”“知识元”等细粒度检索结果^[14]、一站式检索方式^[15], 以及语义相关度较高的检索结果^[14]。信息的组织方式通常可直接反应信息检索和信息服务的效果, 当前普遍使用的DC、MARC等元数据标准主要以粗粒度的信息资源为单位(一本书、一个网页等), 缺少对细粒度单元(章节、段落、句子等)的描述标准, 无法真正满足图书馆资源组织的迫切需求, 而针对用户个性化需求的多粒度知识服务更加匮乏。

因此,针对图书馆服务现状,本文针对当前用户的多粒度精准服务需求,提出基于多粒度的图书馆知识服务创新模型。

2 多粒度的概念与内涵

粒度的概念由Zadeh^[16]于1997年在其模糊信息粒化和词计算理论中提出,并明确粒化、组织和因果三个基本概念。粒化指将整体分解为部分;组织是粒化的逆过程,即将部分合并为整体;因果是部分间的因果关联。根据粒化的分解程度(即粒度),可分为粗粒度(最大粒度)、细粒度(基本粒度,不可再分的粒度)、中粒度(介于粗粒度和细粒度间,可以有不同分解程度的多粒度)三个层次^[17]。多粒度指不同粒度(细粒度、中粒度、粗粒度)共存的现状。本文提出的多粒度概念主要体现在用户需求的多粒度和资源组织的多粒度两个维度。用户需求的多粒度主要区别于图书馆用户传统的文献查新、文献检索等以文献为单元的粗粒度信息服务需求;资源组织的多粒度区别于当前图书馆信息组织过程中单一粒度的文献、文件、书目等,是用户需求多粒度作用后的结果。

2.1 用户需求的多粒度

随着大数据环境下信息冗余、信息迷航、信息焦虑等问题日益加剧,用户的信息服务需求不再是以文献、网页等为单位的单一信息传递或推送服务,而是面向问题解决方案的不同层次的知识服务。从当前图书馆的核心竞争对手网络搜索引擎看,大众越来越偏向细粒度的精准信息服务,而学习型用户偏向以文献为单元的粗粒度信息服务。

2.2 资源组织的多粒度

要满足用户的多粒度需求,首要任务是改善当前仍以文献为单一粒度的信息组织方式,即控制信息组织的粒度,使其从文献细化到知识,从而实现多样化粒度的组织方式。具体到资源内容的组织过程,即实现对文献信息中更小粒度知识点的抽取、标注、揭示和聚合。细粒度知识点通常指文献中包含的概念、公式、图表或数据,即当前在精准信息组织中提到的“知识元”^[18]。知识元是知识组织的基元,是构成知识结构的最小独

立单元,用来表示针对特定问题的解决方案,可以是概念、方法、规则、公理等数据或事实,以及实例化的知识。因此,细粒度的知识组织要抽取文献信息中的知识元,建立以知识元为核心的知识元关联及知识库。

3 基于多粒度的图书馆知识服务模型

面对用户日益精准与个性化的多粒度、语义化信息需求,本文提出基于多粒度的图书馆知识服务模型。如图1所示,整个模型分为馆藏资源层、知识组织层、知识聚合层及知识服务层。馆藏资源层主要包含图书馆各类型的实物版(如期刊文献、书目著作、报纸、报告、专利等)及电子版资源(如pdf、音频、视频、图片等),其他依次是基于多粒度的知识组织层、基于语义的知识聚合层及知识服务层。

3.1 知识组织层

资源的组织方式直接决定了信息的检索方式及信息服务的效果,传统的信息组织方式是基于文档篇章粗粒度的单一组织方式,即对整个文件或整篇文献进行标注,而未深入文档内部进行标注。虽然全文检索实现了对全文内容的索引,但基于单汉字或语词的索引方式并未表现出文档内部结构间存在的逻辑关系。

依照多粒度的概念,可将图书馆馆藏资源的粒度划分为细粒度、中粒度、粗粒度三种。以科技期刊为例,三种粒度分别具体表现为以句子为单元的知识元、以小节为单元的知识单元和以篇章为单元的知识群。知识元作为最小粒度的信息,多代表一个概念、一个公式、一个结论等,其能解决某一简单问题并给出直接答案;知识单元属于中粒度的信息,通常指文献中的一个小节,是相关知识元相互关联的集成,可以解决复杂的用户问题;知识群属于粗粒度的信息,多指一篇完整的文献或书目,是相关知识单元的汇聚,可以解决用户较复杂的问题。图2为本文提出的多粒度资源组织的逻辑模型。

基于多粒度知识组织的核心思想是实现文献内容结构化、多层次、多粒度的组织,以一篇期刊论文为例,主要过程分为3个步骤。

3.1.1 文本结构分析

期刊文献由物理结构和逻辑结构两部分组成^[19]。

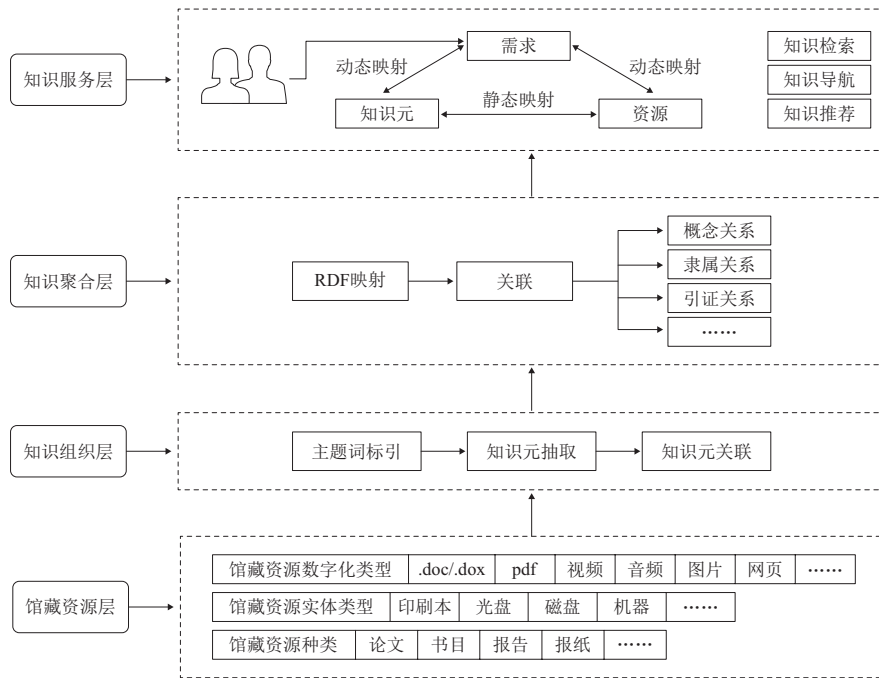


图1 基于多粒度的图书馆馆藏资源集成知识服务模式

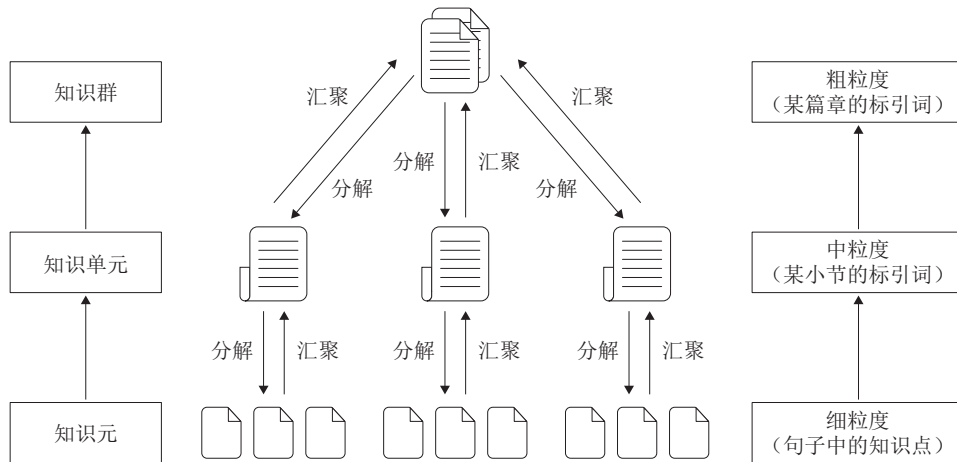


图2 基于多粒度的图书馆馆藏资源组织模型

文献的物理结构包括题目、作者、摘要、篇章、段落、句子、词语、引文，针对文献的物理结构信息可较容易地对整篇文章的作者、机构等元数据信息进行标引；逻辑结构信息包括标题、层次、段落、句子、主题词、标识词、中图分类号，因此，对文献多粒度的标引即根据文献的逻辑结构对其篇章、小节、段落等逐一以主题词进行标引。此外，在本文的多粒度知识组织过程中，还须对文献段落间的关系、句子间的关系、上下文及位置进行记录，为下一步抽取知识元的地址信息作铺垫。

3.1.2 基于多粒度的主题词标引

在主题词多粒度的提取过程中，本文将一篇文献分为三个粒度，即整篇文献、一级标题、二级标题。其中，整篇文献的标引为粗粒度，一级标题和二级标题下的段落为中粒度，二级标题下文本内容所包含的知识点，即抽取的知识元为细粒度。主题词的标引旨在解决文献资源粗粒度和中粒度的标引工作。针对不同粒度（粗粒度或中粒度）的文本信息，以该粒度为单位，将该粒度下的文本块进行分词、词性标注、去停用词、计

算TF-IDF权重,最后确定该文本块的候选主题词。

3.1.3 构建知识库

知识元是知识库构建的基元,知识元的抽取是实现资源细粒度检索的开始。针对结构化期刊文献知识元,本文采用一个七元组来表示,即 $K = \{ \text{编号, 导航, 地址, 特征词, 属性, 属性值, 内容} \}$ 。

①编号:是针对知识元所采用的一种编码符号,主要用于知识元的标识。②导航:主要标识该知识元在期刊文献中的逻辑结构信息,即具体的位置信息。③地址:指该知识元所在期刊的存储信息。④特征词:指针对知识元内容提取出的具有一定描述知识元信息能力的词语或短语,通常是领域本体库中的相关概念。⑤属性:特征词的谓语动词,可称为关系。⑥属性值:特征词的宾语,可称为客体。

本文抽取的知识元并非针对整篇文章的全部内容,而是针对第二步提取出的粗粒度、中粒度的主题词所包含的知识内容,即将粗粒度、中粒度的主题词作为知识元的特征词,并以特征词为向导在其对应的文本位置提取主题句。然后,对主题句进行分词、词性标注

等初步处理,为进行依存句法语义分析提供基础。依存句法理论能对句子进行形式化的句法和语义分析,使计算机能以依存树的形式显式地界定句子成分(词或者短语)间的依存关系,实现对语义的理解,为知识元抽取提供较好的理论基础和技术手段^[20]。如图3例句所示,知识元是最小的知识元素,是构造知识系统的基元。依据其中的SBV(主谓关系)、VOB(动宾关系)即可提取相应的谓语和宾语。

3.2 知识聚合层

面对资源异构带来的“信息过载”和“信息孤岛”问题,图书馆紧随时代信息技术前沿,不断完善知识服务环境并进行模式创新,从Web环境下基于OPAC系统的资源整合、基于资源导航的资源整合、基于跨库检索系统的资源整合到如今基于语义的资源深度聚合^[21],图书馆逐步从数据到信息再到知识聚合的方向发展。目前使用广泛的基于跨库检索系统的资源整合方式,虽较好解决了前端的统一检索,但在用户实际使用过程中,由于信息资源内部语义关联性较差,因此检索效果依旧不够理想。

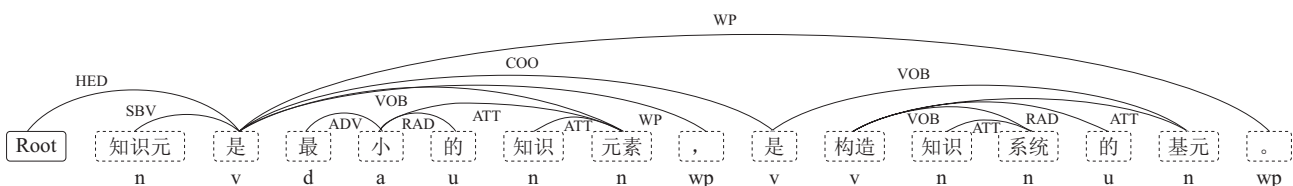


图3 依存句法分析示例

注:下方标识为词语的词性标注,上方标识为依存句法分析中语句成分间的关系,具体含义参见哈尔滨工业大学语言技术平台。

知识聚合指通过一定技术手段(如统计分析、数据挖掘、文本分析等),对可能存在的隐性关联知识单元进行凝聚,以提取知识单元间的内在关联为手段,构建多维、多层、互相关联的知识体系^[22]。随着语义网技术的不断发展,关联数据成为当前语义网实现过程中的最佳实践,为知识聚合提供有效途径。关联数据采用三元组<主语,谓语,宾语>作为基本数据模型。知识组织层提取出的知识元特征词、属性和属性值分别映射为关联数据三元组<主语,谓语,宾语>,其中宾语可对应属性值或新知识元,从而将不同的知识元相关联。知识聚合主要利用关联数据对馆藏资源间的概念关系、隶属关系、引证关系、映射关系等进行揭示,以实现基于语义的关联。

3.3 知识服务层

随着用户信息需求层次不断提升,由图书馆单一粒度的信息组织返回的检索结果已无法满足用户的个性化需求,非学习型用户在查询日常简单问题时,常倾向于使用网络搜索引擎。仅当用户在学习和研究过程中需要大量系统、完整、严谨的文献或文件时,才会转向使用图书馆服务。

多粒度馆藏资源分层组织有效地解决了用户获取各种粒度的信息或知识问题,但如何根据用户的检索提问和搜索行为,推送给用户最合适粒度的信息资源仍是图书馆知识服务的焦点。直接的方式是在用户检索过程中,让用户根据自身信息需求选择期望的信息

粒度,当用户想要了解更多所选粒度的信息时,可进一步了解。但更智能化、个性化的服务仍依赖于对用户行为 and 兴趣的建模,实现用户需求、资源和知识间的映射关系。

4 基于多粒度的图书馆知识服务功能

4.1 基于多粒度的一站式知识检索服务

当前图书馆馆藏数据资源基本处于分布式存储、分布式访问的状况,读者在检索过程中需要对各数据库逐一访问,同时基于关键词的检索方式与匹配机制对用户的关键词提取能力及语义扩展能力要求较高,否则返回的结果通常语义关联性较差,导致大量冗余信息以粗粒度的文件或网页形式展现给用户。因此,知识检索服务要解决的两个核心问题是基于多粒度的语义检索结果和一站式的检索环境。

相较传统检索方式仅给出文献链接列表信息,基于多粒度的一站式知识检索服务通过多粒度知识组织方式和基于语义的知识聚合方式,将大幅降低用户信息检索成本并减弱对用户搜索能力专业性的要求。用户以自然语言表达自身检索需求,检索系统通过一定规则的需求描述模型来抽取用户需求概念,建立用户需求概念与知识元间的映射,从而聚集相关资源内容,并返还给用户知识元(细粒度)、知识单元(中粒度)、知识集群(粗粒度)三种粒度层次的知识,使用户获得所需的知识服务。

4.2 基于可视化展示的知识导航服务

语义导航是一种基于本体网络的馆藏资源知识结构及其分布的可视化展示方法,能有效地描述资源间的语义关系^[23]。基于可视化展示的图书馆知识导航服务将对检索结果从结果聚类、结果关联两个维度进行全方位展示。结果聚类指应用文档聚类技术,将文献检索结果分成若干称为“簇”的子集,每个簇中的文献之间具有较大的相似性,而簇之间的文献具有较小的相似性^[24];结果关联指对检索结果按照研究时间、研究领域、研究地域等进行关联可视化,形成以检索对象为中心的知识图谱,并得出相应的研究趋势。结果聚类提高了检索结果的查准率,结果关联得到检索结果的知识图谱,二者相辅相成,旨在全面提高用户的满意度。

4.3 基于用户动态需求的知识推荐服务

基于用户动态需求的知识推荐依赖两个方面:一是根据用户的浏览痕迹、浏览行为,与用户数据库中其他用户的相似行为进行匹配,并给出相应的关联推荐;二是借助语义匹配的方式,将用户具体某次的检索需求映射为本体概念,然后与知识库中的全局集成本体进行动态匹配,并将与之关联的资源推荐给用户。在此过程中,用户需求模型的表达是关键,用户需求通常分为用户检索输入的显性信息需求,以及通过其日常检索行为特征、职业、兴趣等表现出的隐性信息需求。

5 结语

馆藏资源的多粒度精细化组织方式及资源间的语义聚合,一直是图书馆知识服务工作中知识组织、知识发现的研究重点,因此将二者结合,应用到图书馆今后的知识服务工作是必然趋势。本文基于提升和完善图书馆知识服务效果的目标,提出基于多粒度的图书馆知识服务模型,并给出基于多粒度知识服务的具体功能。本文对今后图书馆的知识服务开展方向有较强的理论指导意义,但在具体应用实践上有待进一步加强。

参考文献

- [1] KWANYA T, STILWELL C, UNDERWOOD P G. Library 3.0: Intelligent libraries and apomediation [M]. Amsterdam: Elsevier, 2014: 35-45.
- [2] KWANYA T, STILWELL C, UNDERWOOD P G. Intelligent libraries and apomediators: Distinguishing between library 3.0 and library 2.0 [J]. Journal of Librarianship and Information Science, 2012, 45 (3): 187-197.
- [3] CASEYM E, SAVASTINUKL C. Library 2.0: a guide to participatory library service [M]. Information Today/Medford: Information Today, 2007: 5-7.
- [4] CHAD K, MILLER P. Do libraries matter? The rise of Library 2.0 [R/OL]. [2017-12-01]. https://www.immagic.com/eLibrary/ARCHIVES/GENERAL/TALIS_UK/T051111C.pdf.
- [5] STEPHENS M. Tame the web: libraries and technology [J]. Retrieved January, 2007 (3): 2008.
- [6] MILLER P. Library 2.0: The challenge of disruptive innovation [J]. Library, 2006 (2): 11-17.

- [7] MANESS J M. Library 2.0 theory: Web 2.0 and its implications for libraries [J]. Webology, 2006, 3 (2): 1-12.
- [8] IrisJastram. Library 3.0? [EB/OL]. [2017-08-13]. <http://lonewolf librarian.wordpress.com/2008/08/15/library-30081508/>.
- [9] 张晓林. 走向知识服务: 寻找新世纪图书情报工作的生长点 [J]. 中国图书馆学报, 2000, 26 (5): 32-37.
- [10] BELLING A, RHODES A, SMITH J, et al. Exploring library 3.0 and beyond [R]. Victoria: State Library of Victoria, 2011.
- [11] 刘佳. 数字图书馆知识服务能力评价研究 [D]. 长春: 吉林大学, 2010.
- [12] CLAIR G T. Knowledge services: your company, skey to performance excellence [J]. Information Outlook, 2001, 5 (6): 6-8.
- [13] 李晓鹏, 颜端武, 陈祖香. 国内外知识服务研究现状、趋势与主要学术观点 [J]. 图书情报工作, 2010, 54 (6): 107-111.
- [14] 索传军. 网络信息资源组织研究的新视角 [J]. 图书情报工作, 2013 (7): 5-12.
- [15] 张海涛, 宋拓, 刘健. 高校图书馆一站式知识服务模式研究 [J]. 情报科学, 2014 (6): 104-108, 113.
- [16] ZADEH L A. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic [J]. Fuzzy Sets & Systems, 1997, 90 (90): 111-127.
- [17] 冯儒佳, 王忠义, 王艳凤, 等. 科技论文的多粒度知识组织框架研究 [J]. 情报科学, 2016, 34 (12): 46-50.
- [18] 温有奎, 徐国华, 赖伯年, 等. 知识元挖掘 [M]. 西安: 西安电子科技大学出版社, 2005: 143-144.
- [19] 温有奎. 基于“知识元”的知识组织与检索 [J]. 计算机工程与应用, 2005, 41 (1): 55-57.
- [20] 车万翔, 张梅山, 刘挺. 基于主动学习的中文依存句法分析 [J]. 中文信息学报, 2012, 26 (2): 18-22.
- [21] 毕强, 牟冬梅. 语义网格环境下数字图书馆知识组织理论、方法及其过程研究 [J]. 图书情报工作, 2007, 51 (8): 6-9.
- [22] 贯君, 毕强, 赵夷平. 基于关联数据的知识聚合与发现研究进展 [J]. 情报资料工作, 2015 (3): 15-21.
- [23] 王忠义, 夏立新, 王伟军. 云环境下数字图书馆知识管理研究 [J]. 情报科学, 2015 (3): 13-17.
- [24] 韩家炜, 坎伯. 数据挖掘: 概念与技术 [M]. 北京: 机械工业出版社, 2001: 232-233.

作者简介

陈燕方, 女, 1992年生, 博士研究生, 研究方向: 信息构建、信息服务, E-mail: cyf@ruc.edu.cn。

Library Knowledge Service Innovation Based on Multi-Granularity

CHEN YanFang

(School of Information Resource Management, Renmin University of China, Beijing 100872, China)

Abstract: There are two main problems faced by library, namely the single way of information retrieval results and poor semantic relationship among resource in library, and users' various and individual demands, thus, library knowledge service model should be improved greatly to tackle these problems. Through literature review about knowledge service condition in library currently, this paper, firstly, introduces the concept of multi-granularity, and then proposes the library knowledge service model based on multi-granularity, including four layers, library resource, knowledge organization, knowledge aggregation and knowledge service. Finally, this paper proposes three specific functions of library knowledge service model based on multi-granularity.

Keywords: Library; Multi-Granularity; Knowledge Organization; Knowledge Service; Semantic

(收稿日期: 2017-12-14)