

新场景、新技术背景下领域知识组织与服务的创新

——评《网络社区知识聚合与服务研究》

领域知识组织与服务是信息资源管理学科的核心课题。近年来,有两个该课题的研究动态值得关注:其一其场景逐步向网络社区延伸,用户基于特定兴趣或目标聚集于诸多网络社区并积累了大量领域知识资源,由此突出了网络社区场景下知识资源组织和服务的重要性。围绕这一问题,主要有“用户”和“资源”两个角度的研究:前者侧重探索网络社区中用户行为特点和影响因素,数量众多;后者侧重对网络社区中的知识资源开展挖掘和重组,这类研究则相对不足。进一步而言,“服务”层面研究更为欠缺。其二是随着以深度学习为代表的计算机技术快速发展,自然语言处理算法模型在执行领域知识分析任务上不断取得突破,领域知识组织方法逐步从传统的分类、聚类、标签标引、主题分析方法深入到基于领域知识结构的方法,典型的是在构建领域知识图谱的基础上开展知识组织。

基于以上两个研究动态,研究者开始探索网络社区场景下的领域知识组织方法并开展知识服务。这既是在新场景下对信息资源管理学科与生俱来的知识组织优势的检验,又是在新技术环境下优化传统知识组织方法、知识服务模式的必然要求。《网络社区知识聚合与服务研究》一书正是此方面的代表之作,是新场景、新技术背景下对领域知识组织与服务的创新,该书有以下3个方面的特色。

1 网络社区新场景下知识组织理论的深化

网络社区场景下的知识组织具有“面向用户”和“基于用户”的双重要求,涉及词汇、文本、用户等多粒度知识单元及其间多种类型的知识关联,因而对知识组织理论提出了新的要求。该书以领域知识分析为理论视角,在开展网络社区知识聚合时,以词汇为核心知识单元,在充分挖掘词汇语义及其语义关联的基础上,有效关联网络社区中的用户、文本,并构建出相应的知识聚

合模式。为了落实这一核心目标,该书在理论上作了3个方面的深化探索。

其一,在知识单元层次上,围绕词汇语义挖掘核心目标,该书提出“词汇语义链”理论。根据词汇蕴含的领域语义内涵,该书将词汇按链式结构划分为“词语→领域术语→领域实体→领域概念”4个语义阶段,并系统地归纳了领域知识分析视角下词汇语义阶段提升所需的背景信息。“词汇语义链”理论很好地解决了当前知识组织理论与实践词汇对象的界定混淆问题。其二,在知识关联层次上,该书提出并完整地论证了领域多元概念关联体系融合理论。该书将领域知识关联归纳为领域结构化概念关联体系和概念相似度体系,涵盖了基于领域知识结构的知识组织体系和基于统计挖掘的知识关联体系。该书论证了基于优势互补原则的领域多元概念关联体系融合的可行性、主体选择问题和融合方式。以此为基础,该书进一步探索网络社区中“领域实体-主题-文档”和“用户-领域实体-资源”两种多粒度知识关联体系。其三,在知识聚合模式层面,该书在挖掘网络社区知识单元和知识关联的基础上,将传统的“领域知识基于什么来聚合”问题进一步深化到“领域知识聚合成什么形式”上。该书结合网络社区用户探索和利用知识过程中用户、文本、词语的切换路径,以多元知识关联体系为聚合依据,提出了包括分面式导航、知识元链接、语义检索、潜在关联文本发现等的知识共聚模式。

2 围绕领域泛化能力提升的知识组织技术优化

近年来,基于深度学习的自然语言处理技术快速发展,以领域实体识别和实体关系抽取为核心的知识图谱构建技术有力地促进了知识组织的开展。然而,这些技术方法在应用于具体领域时大多严重依赖人工标注样本。人们似乎已有共识:如果某领域有很好的人工

标注语料,那么知识抽取、本体或知识图谱构建等技术工作将能顺利开展。在实践中,制约特定领域知识组织开展的更为急迫的问题,往往不是“不知道采用哪些技术方法”,而是“难以有效地人工标注出一定规模的领域语料”来使用相应技术方法。

网络社区中的知识内容具有明显的专业领域性,因而该书十分注重解决这一问题。该书充分吸收了近几年自然语言处理领域新兴的“预训练+微调”“小样本学习”技术方案,围绕领域泛化能力提升这一目标,提出一系列方法以解决网络社区知识组织面临的低资源问题。例如,针对网络社区中领域术语频次较低导致词汇语义学习不充分的问题,该书不仅采用了“通用预训练词向量+领域小规模语料训练微调”的方案,还进一步结合领域术语的构词法特点,将领域术语拆分为词素,从通用预训练词向量中寻找词素语义向量来拼接形成领域术语初始向量,有效地克服了低频术语语义学习受初始随机向量影响的问题。针对领域实体关系抽取需要的人工语料标注成本过高问题,该书提出了基于词嵌入类比的网络社区领域实体关系识别方案,充分发挥领域已有知识库“量少质高”和领域未标注语料“规模效应”的特点,利用小规模知识库开展对大规模未标注语料的知识挖掘。可以说,该书提出并验证的技术方案很好地解决了知识组织面临的领域泛化难题,具有很好的实践应用价值。

3 系统涵盖“用户需求-内容挖掘-知识组织-知识服务”全流程

该书的内容组织较为完善,完整地涵盖了网络社区

知识组织与服务的全过程,且其中核心章节均有相应的实验或实证支撑,因而各种知识组织实践工作都能在该书中找到参考借鉴。

首先,该书从网络社区中的用户和知识交流模式出发,提出了面向用户的知识组织与知识服务需求,进一步结合网络社区中的知识聚合对象,提出了面向网络社区的知识聚合理论模型。其次,围绕网络社区中的领域知识挖掘问题,递进式地完成领域知识单元挖掘、领域语义知识关联挖掘、多粒度多元关联挖掘3个子任务。在以领域知识挖掘结果为知识组织依托的基础上,该书进一步提出领域分析视角下的网络社区知识聚合模式及其实现方式,解决了“依据领域知识将网络社区资源聚合成何种形式”的问题。最后,该书分别围绕知识推送服务、知识发现与知识创新服务两大类知识服务需求,探索了网络社区中基于知识聚合的知识服务模式及其实现方式,并结合案例开展了分析。

4 结语

总体而言,在新场景、新技术背景下,作为信息资源管理学科核心课题的领域知识组织与服务研究、实践正处于转折期,《网络社区知识聚合与服务研究》一书可谓是承前启后之作。该书不仅系统地归纳吸收了信息资源管理学科在知识组织与服务理论方法上的传统优势,还突出了对当前新场景、新技术环境下领域知识组织与服务理论方法的探索与创新,对未来的知识组织与服务研究与实践具有很好的启发意义和借鉴作用。

章成志

南京理工大学经济管理学院信息管理系