

## 首届“数据分析与知识发现”学术研讨会在京召开

2017年6月8—9日,首届“数据分析与知识发现”学术研讨会暨“面向决策的大数据分析”培训班在北京世纪金源香山商旅酒店举办。会议由中国科学院文献情报中心主办,中国科学院文献情报中心科技期刊与知识服务中心和《数据分析与知识发现》(原《现代图书情报技术》,2017年更名)编辑部联合承办。会议邀请了国内外从事相关研究的24位知名专家学者做报告,共有来自全国18个省市的大学和科研机构的266人参加了会议。

专家报告共分为7大部分。第1部分的报告包括“大数据驱动的科学发现:愿景与挑战”“大数据驱动的图书馆智能管理与服务系统”“知识图谱与事件学习”。中国科学院计算机网络信息中心的黎建辉教授剖析了科学大数据的本质,以及大数据驱动科学发现的特点、技术需求,他指出,目前数据开放共享面临着3大挑战:一是如何从数据源变为数据资源,二是如何驱动数据的快速流动,三是算法和模型的选择。中国科学院软件研究所的王永吉教授阐述了大数据驱动的图书馆智能管理服务的方法、技术和应用,并针对图书馆的业务需求,提出了数据存储、服务探索、管理决策和人才培养等多方面的策略和建议。清华大学计算机科学与技术系李涓子教授概述了知识图谱和知识共享的基础理论,并介绍了其在事件知识学习上的应用与研究。第2部分的报告有“学术大数据的分析和挖掘”“教育大数据的运用——基于内容管理和服务的视角”“IT技术持续发展及其对教育的影响”。大连理工大学软件学院的夏锋教授以学术年龄对合作模式的影响为例,说明了分析和挖掘科研大数据的意义。华中师范大学信息管理学院的夏立新教授以实例说明通过构建教育大数据聚合与服务平台,有助于解决教学过程质量管理、学生学习成长监测、教育管理决策能力不足等问题。北京理工大学计算机学院的牛振东教授重点介绍了人工智能、大数据等IT技术的发展并探讨了其对智能数字图书馆及网络教育发展的影响。第3部分的报告包括“测度技术融合的数据分析方法研究进展”与“‘AI+慢性病管理’使精准医疗成为可能”。中国科学院文献情报中心方曙研究员对基于专利数据测度技术融合程度与趋势的现有分析方法进行了较系统的介绍。大连理工大学的金博副教授提出采用机器学习和医疗大数据分析的方法,在前期积累的海量医疗数据基础上,构建人工智能+慢性病管理的模式,以数据为驱动,使精准医疗成为可能。第4部分的报告包括“关联数据的消费:查询、分析与可视化”“数据驱动的用户调研问卷自动生成研究”“基于主题模型的无监督文本分类技术”“领域科技论文主题研究脉络挖掘研究与实践”。南京大学信息管理学院欧石燕教授重点讲述了一种面向图书馆关联数据的自动问答新方法。南京理工大学经济管理学院章成志教授以亚马逊网站商品的用户评价为例,介绍了自动问卷调研技术在中美用户购物行为研究中的应用。武汉大学计算机学院李晨亮副教授介绍了一种基于主题模型的无监督文本分类技术。该技术通过类别语义相关的少量种子词,将单次共生关系和概率主题模型相结合,从而达到文本分类的目标。中国科学院文献情报中心知识技术研发中心的乐小虬研究馆员以气候变化领域研究论文数据为例,分析了科技论文组织结构特征及主题与句子间的上下文信息,综合利用深度学习及半监督机器学习模型对主题的研究线路、层级结构及核心主题句进行了识别和汇聚。第5部分的报告有“Dependencies for Graphs”“大元数据分析及其对科学学与科学政策的意义”“大数据与智慧城市”。来自英国爱丁堡大学信息学院的樊文飞教授介绍了名为图实体依赖(GEDs)的图依赖关系,它是图模式和属性依赖关系的组合。重点讲述了从特征描述到复杂实体的可靠性、含义、验证和有限的公理性,并展示GEDs如何帮助进行不一致检测和知识提取。美国雪城大学信息学院的秦健教授从研究了4年的项目中获得的经验和教训出发,依次阐述了在数据采集、预处理、质量控制以及研究设计、性能评估、联合文档协作等过程中的相关概念和计算工作。清华大学信息技术研究院邢春晓教授介绍了大数据驱动下的新型智慧城市研发现状,并以心血管疾病大数据管理平台和智慧养老为案例进行了说明。第6部分包含“信息哲学视域下的数据驱动的科学知识发现方法论思考”“面向同行评议的专家社会关系研究”“数据驱动的科学知识老化模式识别:‘睡美人’与‘昙花一现’现象”“层次知识网络及其骨架结构的稳健性——基于混沌和分形理论的视角”共4场报告。南京理工大学信息管理系王曰芬教授从信息哲学视角,研究数据驱动的科学知识发现方法的目标、核心理论基础、特征、组成要素与应用范围等。中国科学技术信息研究所袁军鹏研究员提出将评议人与被评议人之间的关系分析识别并根据紧密

程度予以合理量化,在评审中,将量化后的社会关系回避处理,或者予以不同权重设置,从而保证评议结果的公信力。浙江大学信息资源管理系李江副教授介绍了科学论文中的“睡美人”和“昙花一现”现象,提出如何评估低被引和零被引论文的价值和如何识别这两种模式的问题。东北师范大学信息科学与技术学院腾广青教授从分形学的视角对基于关联频次提取频次知识网络的理论基础进行了阐释,并验证了层次知识网络与原始知识网络具有等效性。会议最后一部分共包含5场报告,分别为“机器智能与数据服务”“基于大数据技术的可配置计算型情报分析”“基因大数据驱动的农作物精确育种关键技术”“关联大数据的技术挑战、对策与实践”以及“建设基于大数据计算的新型知识服务模式”。上海图书馆副馆长刘炜研究员提出,当今的数据服务应引入人工智能方法,充分利用机器学习的技术和能力,突破传统思维,把图书馆情报服务提升到新的发展阶段。北京大学信息管理系主任李广建教授认为,大数据环境下,情报分析正朝着定量化、自动化、可重复、可机器处理的方向发展。介绍了基于大数据技术的可配置计算型情报分析系统。中国农业科学院农业信息研究所所长孙坦研究员分享了本所设计的覆盖农作物精确育种大数据资源建设与治理、育种大数据基础设施建设、模型研制与平台开发、关联分析与可视化分析的完整技术和实施路线。中国科学院计算机网络信息中心大数据应用服务技术实验室沈志宏主任针对大规模、动态迭代式再组织的关联数据带来的存储和高效加工处理问题,分析其挑战,提出了解决方案。中国科学院文献情报中心副主任张智雄研究员以中国科学院文献情报系统“十三五”重点突破方向“分布式大数据知识资源体系”为基础,重点论述了中国科学院文献情报系统分布式大数据知识资源中心及其相关服务体系建设、大数据治理框架体系建设等主要思路、举措和进展。

中国科学院文献情报中心张晓林教授做了会议总结。本次会议以全景式的深层扫描,聚焦大数据驱动的数据挖掘、语义计算、内容分析、知识发现、智能决策等方面的技术、方法、系统和机制研究,既注重理论研究的深度,也探讨了应用领域的广度,为该领域的跨学科研究与交流提供了很好的平台。

(关志英)

## 《河南科技学院学报》“图书馆史研究”专栏 征稿启事

《河南科技学院学报》(社会科学版)是由河南省教育厅主管、河南科技学院主办的国内外公开发行的学术刊物。为加强特色栏目建设,进一步提升研究水平和学术影响,学报编辑部特邀河南师范大学图书馆从2014年第1期开始,创办“图书馆史研究”专栏,单月出版,每年6期。专栏立足繁荣学术,免收版面费,录用稿件根据学术质量,优稿优酬。

诚挚欢迎图书馆界各位学人不吝赐稿!

### 一、稿件内容

- (一)国内外图书馆史研究的现状、意义与反思;
- (二)图书馆史基础理论与基本问题研究;
- (三)图书馆史史料整理及相关研究;
- (四)中国图书馆史研究;
- (五)外国图书馆史研究及中外图书馆史比较研究;
- (六)图书馆史与其他学科的交叉研究等。

### 二、联系方式

投稿邮箱:hnkjxyxb@126.com

联系电话:13598671251(郑爽);13937340212(胡卫玲);18637376638(苏全有)

联系地址:河南省新乡市建设东路46号 河南师范大学图书馆 邮编:453007