

“ICSTI 2011” 知识组织与知识发现分论坛综述

□ 李颖 王莉 乔晓东 / 中国科学技术信息研究所 北京 100038

摘要: “知识组织与知识发现”是实现“迈向知识服务”ICSTI 2011大会理念的技术基础。该论坛录用了31篇来自国内外高质量研究论文,11名研究学者从不同角度,发表了其本人或所代表机构的研究成果。文章在简要总结分论坛总体情况的基础上,重点论述这11篇会议演讲的主题,最后,作出要点的分析与总结。

关键词: 国际科技信息年会2011, 知识组织, 知识发现, 知识服务

DOI: 10.3772/j.issn.1673-2286.2011.08.004

1 “知识组织与知识发现”分论坛的概要

科技信息机构承载着信息资源服务的重要职责,在数字环境下,科技信息资源的服务正从单纯的资源传递步入知识服务。本届大会的理念“迈向知识服务”,也体现了全球创新趋势中,科技信息机构的历史使命。知识服务的技术基础为知识组织与知识发现,这是实现知识服务的重要研究课题。为此,本届会议设置了“知识组织与知识发现”分论坛。

本论坛由ISTIC信息技术支持中心负责。近年,信息技术支持中心在研究知识组织与发现、开发相关知识服务系统方面,取得了跨越式的进步。在开放资源、开放办所的方针下,中心与国内外相关机构建立了多边的合作交流关系。为促进更为广泛、深入的交流与合作平台,中心邀请了国内外相关研究机构和个人参与论坛的讨论。基于组委会的安排,论坛指定了来自美

国、比利时、日本及国内学者11人进行口头演讲,31篇优选全文在会议录和大会的官方网站上进行了公开出版。由于本论坛涉及的议题广泛,参与度高,晚到会者甚至站立着聆听了发表者的演讲。

本文重点论述11名演讲者的主题,有意了解全文的读者,可以通过本文提供的线索,参考相应的全文^[1-2]。

2 “知识组织与知识发现”分论坛的演讲题目和内容

本论坛特邀国家科技图书文献中心NSTL副主任吴波尔女士、中国图书馆学会学术委员会委员、中国图书馆学会数字图书馆专业委员会副主任、中科院国家科学图书馆馆长张晓林先生,以及加拿大国家研究委员会科技信息研究所总干事Pam Bjornson教授(女士)(分会主席)主持讨论。

7日,主持为吴波尔和Pam

Bjornson两位女士,会议演讲题目如下:

■ Ronald Rousseau, 比利时鲁汶大学

题目: 数据挖掘、知识表现和引文网络——金三角

内容: Ronald Rousseau首先图示了数据挖掘的过程,强调信息科学的核心工作和现代图书馆的工作为知识表现,网络计量学是数据挖掘的结果,所有引文研究为处理表现问题。引文网络无处不在、引文网络中的知识发现及网络的各种应用等等。Ronald Rousseau的演讲让我们对信息科学的“金三角”有了整体有机的认识。

■ 陈超美, 美国德雷塞尔大学

题目: 创新的本质: 机制、计量及分析

内容: 突破性、创新和变革性

研究常被认为非常有创意、高风险和不可预测。在本演讲中,陈超美教授讨论了创新的一些基本问题。论及了什么是创造的本质、是否有通用机制、是否有证据可以帮助我们早期阶段认识到潜在的变革性研究。该演讲介绍了创新的理论框架,引出信息计量学,它可以用来计量科学领域中一个想法或新建议的创新潜力。最后,演示了一些初步的结果,显示了在个人、机构和社区的不同层面,在情境意识和竞争情报方面,如何将这一方法应用到基于证据的分析和决策领域。

■ Mari Jibu (治部真理), 日本科学技术振兴机构 (JST)

题目: 从“科学之科学”到“科学政策之科学”

内容: 本演讲首先介绍了与研究有关的背景,涉及科学计量学和文献计量学、科学计量学和文献计量学的整合、科学地图、政策地图、J-GLOBAL (JST科学技术综合链接中心) 远景设计等事项。介绍基于汤森路透公司的Web of Science和Derwent Innovation Index、爱思唯尔公司的Scopus Custom Data和EPO PATSTAT等数据库,整合论文和专利,对JST资助的科研成果,进行了分析,并将结果可视化,最后给出即将实现的新功能。

■ 袁军鹏, 中信所情报方法中心

题目: 国际科技合作对中国科技能力影响分析

内容: 本研究介绍利用科学计量学、数据挖掘技术,以1994-2006

年Web of Science收录的中国发表论文为样本,从论文的机构类型分布、引用影响分布、国际科技合作与全部中国文献引用影响的对比、中国与国际著名机构的合作分布进行了研究,从科学计量学的视角描绘了1994年以来中国参与国际科技合作对中国科技能力的影响以及中国自主创新能力的演变趋势。提出了中国各机构加强国际合作,提高自主创新能力应采取的对策思路。

■ Noriko Kando (神门典子), 日本国立信息学研究所 (NII)

题目: NTCIR中的跨语言及多语言信息获取研究

内容: 神门教授介绍她本人负责的NTCIR (NII Test Collections for Information Retrieval and Access Systems, 日本国立信息学研究所信息检索和获取系统测试集) 的整体情况、NTCIR中的跨语言和多语言信息获取研究、未来发展方向上的思考。NTCIR是系列化的研讨班,旨在通过大规模测评的基础设施强化信息获取技术领域的研

究。其关键活动是数据集构建、评价方法论的确立,以及论坛的定期举办。数据集的类型有学术资源、新闻、专利及Web资源等,涉及的语种有中文、韩文、日文和英文。主要的任务(即研究领域)为信息检索(跨语言检索、专利、Web、地理信息检索)、QA(单语和跨语言)、自动摘要&趋势&专利地图、观点分析&文本挖掘、社团的研究互动。未来的发展方向:探索性信息获取、通过基于“组件”式的合作研究和开发促进NTCIR的开放发展。

■ Tomoko Steen, 美国国会图书馆

题目: 从STM (Science Technology Medicine) 的数据到知识树

内容: 介绍研究人员的需求从STM数据到全球化复合知识的变化、STM数据的定位和识别、如何将STM数据变成全球化复合知识,需要Lingua Franca (通用语言);需要提供数据与知识之间的桥梁;需要政府的参与和国际合作。她最终强调了在不同语言区域高科技数据



知识组织与知识发现分论坛

获取上的努力,将有助于研究者应用全球化复合知识,强调如何管理指数级增长的数据集的重要性。

■ 夏立新, 华中师范大学信息管理系教授

题目: 基于主题图的电子政务门户研究

内容: 介绍电子政务的现状、主题图的概念和应用、主题图在电子政务领域应用的可行性研究及模块设置(关系型数据库的数据抽取、基于OAI元数据收割、主题图的生成及合并)。

■ 陆伟, 武汉大学信息管理学院教授

题目: 基于域加权聚类算法的网络民意热点话题监测

内容: 介绍了发现新闻中热点话题,可使机构更容易地从新闻中发现关于其自身热点话题。介绍了相关研究、本研究的方法论,即单遍法增量聚类Single-pass Incremental Clustering→域加权→命名实体加权。有关实验,从数据收集、实验过程、实验工具到评价方法,给出了结论,即聚类和域加权的有效性、实体没有像预期那样影响结果。未来的工作,尝试PLSI和LDA等主题模型、增加语义特征计算文献的相似度、对实体影响的注意。

8日,主持为张晓林先生和Pam Bjornson女士,会议演讲题目如下:

■ Boe Horton, ProQuest高级副总裁兼东亚区董事总经理

题目: 把握中国高校生的检索

和发现行为, SummonTM在Web规模知识发现中的角色

内容: 研究动机是掌握中国高校生为取得学位,所需的信息研究和发现、采用的工具手段、影响他们检索行为的因素、明确他们的困境、具有理想功能的发现和检索新工具。研究方法为访谈法,实地考察期间为2011年3月30到4月22日。覆盖城市: 北京、西安、厦门。CALIS所有馆员和学生参与了合作研究。回答机构: 北大、清华、对外经济贸易大学、厦门大学、集美大学、华侨大学、西安交通大学、西安理工大学。通过检索与发现工具的功能,满足简洁、易用、快速的用户需求。ProQuest的SummonTM是Web规模的发现工具,可以在任何地方,快速发现图书馆所持有的最可信资源。该演讲还介绍了SummonTM的特征和中文的可适用性。

■ 林钦佑, 微软亚洲研究院首席研究员

题目: 使用户获得知识——Web检索的下一个前沿

内容: 指出Web研究的重要趋势从信息组织变化为强调用户的信息需求、语义检索、云计算平台和开发者生态系统。语义检索从传统的检索用户输入字面匹配转化为用户的真正意图,当前检索范式的局限性。未来Web检索的前沿通过对用户意图的理解和匹配,使其获得知识、获得创新,是语义检索+社会检索。本研究介绍了微软亚洲研究院的Sempote NeedleSeek项目,它是开放领域语义挖掘和服务:从Web规模资源中发现开放领域语义知识、基于挖掘到的语义知识来回

答用户的提问等。

■ 梁冰, 中信所技术支持中心

题目: 基于NSTL案例研究数字图书馆知识发现框架

内容: 简要介绍NSTL,其文献资源、标准化的 workflows等。介绍NSTL的数字服务平台和文献资源的发现机制,即元数据仓储和检索、基于语境的知识库、跨库检索、接口封装、嵌入式服务、网关服务。描述在“十二五”期间,NSTL基于STKOS(科技领域知识组织体系)的知识发现框架,包括内容建设、内容的深度加工、科技术语服务,以形成完整的知识空间,支持高效便捷的知识服务,插件式数据分析引擎、抽取知识要素表现知识内容、特殊领域知识组织体系、知识服务的场景等。

3 “知识组织与知识发现”要点分析

本论坛基于不同视角,集中论述了有关知识管理和知识发现的课题。演讲者所做的精彩演讲,都很好地反映了开幕式演讲中论及的挑战,我们面对着巨大数量的信息和数据,数据量每18个月翻一番。要特别留意美国国务卿副科技顾问Andrew Reynolds先生关于“黑天鹅”的概念和稀缺不可能事件,其频频出现超出我们的想象,这些事件会改变一切,而我们却经常错过它们,因为在过去我们依据了很多对未来的思考而已。来自王众托院士的演讲似乎是对Andrew Reynolds先生的回应,他发表了系统工程的方法论应用于知识服务和知识可能系

统,既要考虑事实和显性知识,也要考虑整合人的因素和隐性知识。在中信所赵志耘博士的大会演讲中,我们听到了有关决策支持的课题,第二论坛的演讲者很多也论及了此题目。

Ronald Rousseau博士谈到了数据挖掘和知识表现,并能从中获得大量的信息知识。陈超美博士研究了人们如何创新发现及如何识别关联以确立新知识。他采用了“information foraging”(信息寻觅)这一有趣的词汇。分析了以前所谓两个“孤岛”之间的关联,在以前未被关联概念之间,建立了关联。他指出,最重要的新课题会出现在这一位置。治部博士的演讲具有同一理念,用于科学计量学领域。通过日本的J-Global,说明了研究投入与论文和专利数量之间关联的一些观点。发现研究投入的价值是基金机构相当感兴趣的议题。

袁军鹏博士再度使用引用排序来考察源于国际合作的论文增长和影响力。神门典子博士论及了NTCIR面向信息检索的系统方法和专家方法。她指出,需要更多的探索性研究。Tomoko Steen博士说:针对科学知识,我们需要“lingua franca”(通用语言),并为此所

作出的努力:科技教育、英文出版物、专门工作组及全球合作,纵向和横向的国际合作。

夏立新博士指出了电子政务需要进一步的互操作和集成,以便数据可跨省和城市利用。陆伟博士谈到了主题模型和研究案例,分析了在Web上发现非标准信息,比如热点话题和民意,可以扩张,并采用多种形式,但是没有定义结构。

Boe Horton先生概述了对中国大学生的一项最新研究,并不惊奇地发现,他们需要快速、便捷、简单地获取到全方位素材。他指出“Summon”提供了一个类似Google Scholar的Web规模的检索工具,可以发现图书馆所有的最可信资源。把用户从词汇引向所需的主题和意图检索,而非检索你键入的文字。中信所梁冰博士讨论了某种超级叙词表,增加领域叙词表,构建NSTL的元数据集,用于知识系统,达到分面检索。与Boe Horton先生自底向上方法相比,梁博士方法为自顶向下,但都是为了同一目标。

总而言之,该论坛发言者的主要目标,简单来说,是通过智能化工具理解数据或大量信息、检索或提供对过去或未来的见解。引文网

络往往作为资源或有关趋势的情报来源加以研讨,有关研究的影响力及创新研究的发现已经凸显出来。

本论坛非常活跃,同时有很多待解决的课题,听众非常投入。本综述不可能表达这些论文的宽度和深度,请阅读作者的全文,与他们交流,以便更好地理解他们的研究工作。

3 结语

科技创新不仅是中国和谐发展的国策,也是全球共同发展的战略,强化科技信息机构间的国际合作,无疑是重要的举措。ISTIC首次成功举办ICSTI 2011大会,体现了我国在该领域由被动到主动的发展过程,体现了国际合作与交流的重大意义。我们将发挥本次汲取的经验,力争为世界科技信息领域的发展作出更大贡献。

“知识组织与知识发现”这一研究领域,在众多研究者的努力与合作下,必将取得更大的突破,为全球的科技信息用户提供其所需的高层知识服务。在公益服务、OA服务的理念下,未来一定会消除知识发现和获取上的差异,实现理想平等的知识化社会。

参考文献

- [1] ICSTI 2011官方网站[OL]. [2011-06-15]. <http://www.icsti2011.org/eng/index.html>.
- [2] ICSTI 2011资料集[OL]. [2011-06-15]. http://www.icsti2011.org/eng/download_information.html.
- [3] HE DEFANG. The Proceedings of 2011 ICSTI Public Conference [C]. USA: Scientific Research Publishing, 2011.

作者简介

李颖, 博士, 专业: 信息系统。相关研究课题: 知识组织、基于XML的数字出版、基于DOI的文献链接和DRM系统等。E-mail: liying@istic.ac.cn

Review for "Knowledge Management and Knowledge Discovery" Session of ICSTI 2011

Li Ying, Wang Li, Qiao Xiaodong / Institute of Scientific & Technical Information of China, Beijing, 100038

Abstract: "Knowledge organization and knowledge discovery" is supporting technology to achieve the concept of "Upgrading Information to Knowledge" for ICSTI 2011. The forum accepted 31 high-quality research papers from domestic and overseas, and 11 scholars from the different viewpoints, made their speeches. This article briefly summarizes the overall of the session 2, focuses on the themes of the 11 speeches, and finally, gives analysis of the main points.

Keywords: ICSTI 2011, Knowledge organization, Knowledge discovery, Knowledge service