

第22届国际万维网大会

关键词：万维网 社会网络 搜索引擎 数据挖掘

刘奕群 唐 杰
清华大学

会议概况

2013年5月13~17日，第22届国际万维网大会 (the 22nd International World Wide Web Conference, WWW2013) 在巴西里约热内卢举行。随着网络的迅速发展及普及，国际万维网大会自1994年在欧洲核子研究委员会 (Conseil Européen pour la Recherche Nu-

cleaire, CERN) 举办以来，逐渐成为全球学术界、工业界相关领域研究人员交流的盛会。本次国际万维网大会是首次在拉丁美洲举行，也是继 WWW2008 (中国)、WWW2011 (印度) 之后在“金砖”国家举办的第三次会议。

本次大会收到来自50个国家的831篇论文投稿，最终录用125篇 (录用率约15%)。美国 (297篇)、



图1 万维网创始人蒂姆·伯纳斯·李组织论坛

中国(110篇)与德国(54篇)是投稿最多的三个国家,而以色列(39%)、奥地利(24%)与意大利(24%)则是论文录用率最高的三个国家。除了技术论文讨论环节之外,大会还组织了海报展示环节(共录用论文100篇),并同时举办了21场研讨会(workshop),22场辅导报告(tutorial),多个专题论坛(panel)和博士生论坛等。

中国大陆学者中第一作者在本届大会共发表了10篇相关研究论文,作者来自清华大学、北京大学、南京大学、上海交通大学、微软亚洲研究院、中国科学院计算技术研究所、浙江大学、中国科学院自动化研究所、四川大学等高校和科研院所。其中清华大学的研究人员作为第一作者的论文有3篇,来自国防科学技术大学、山东大学等单位的研究人员在海报展示等环节也发表了论文。

大会共吸引了1097名参会者,另外有858人通过在线视频观看了大会报告。

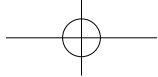
研究论文

与国际信息检索大会(Special Interest Group on Information Retrieval, SIGIR),数据挖掘及知识发现会议(Special Interest Group on Knowledge Discovery and Data Mining, SIGKDD)等专注特定领域的学术会议不同,万维网大会涉及的研究领域遍及网络科学研究的方方面面,学术界与工业界也以极大的热情参与其中。美国斯坦福大学、卡耐基梅隆大学、伊利诺伊大学香槟分校等世界一流大学及谷歌、脸谱、推特、微软等领先企业都发表了众多研究论文。大会研究论文领域分布在一定程度上可以作为相关学术研究方向的晴雨表。

社交网络分析、网络数据挖掘、用户行为分析与个性化、网络信息安全、网络搜索等是会议最热门的主题,投稿量最大,也是录用率较高的。录用论文较多的领域还有互联网商业化、用户交互设计

表1 WWW2013研究论文所涉及的领域分布及与往届会议的比较

领域	投稿量	录用量	投稿比例	录用比例	WWW2008 录用比例	WWW2012 录用比例
行为分析与个性化(behavioral analysis and personalization)	83	17	9.99%	13.60%	/	7.41%
异构数据融合(bringing unstructured and structured data together)	41	3	4.93%	2.40%	/	/
内容分析(content analysis)	59	6	7.10%	4.80%	/	5.56%
互联网商业化(Internet monetization and incentives)	53	10	6.38%	8.00%	7.77%	10.19%
搜索引擎系统及应用(search systems and applications)	78	10	9.39%	8.00%	14.56%	8.33%
互联网安全(security, privacy, trust, and abuse)	82	15	9.87%	12.00%	6.80%	4.63%
语义万维网(semantic Web)	59	7	7.10%	5.60%	8.74%	8.33%
社会网络分析(social networks and graph analysis)	145	22	17.45%	17.60%	8.74%	19.44%
软件体系结构及性能评估(software infrastructure and their performance, scalability and availability)	28	4	3.37%	3.20%	2.91%	5.56%
用户交互与智能终端(user interfaces, human factors, and smart devices)	53	11	6.38%	8.80%	6.80%	4.63%
万维网开发(Web engineering)	55	8	6.62%	6.40%	8.74%	6.48%
网络数据挖掘(Web mining)	95	12	11.43%	9.60%	11.65%	11.11%



领域。与 WWW2008 相比，关于社交网络分析的论文数量和比例有明显上升（增长率超过 100%），这与社交网络产品的广泛兴起有密切关联。另外，与网络信息安全主题相关的论文数量也有较大增长。相比而言，网络数据挖掘、网络搜索等主题的论文比例有一定下降。与 WWW2012 相比，本次大会新开辟了与异构数据融合相关的研究领域，尽管论文数量不多，但仍可以看到大数据研究方兴未艾背景下互联网研究受到的影响。

在社交网络领域，网络结构分析、群体行为分析（包括众包和群体智能）以及网络用户行为建模是研究的热点。总的趋势是网络分析的任务更加细化、微观化。例如，在网络结构分析方面，从网络宏观聚类系数的近似估计，到快速社区发现算法，再到更微观的网络结构洞研究都有不少的研究成果发表。另一方面，将自然语言处理技术融合到社交网络分析中也是一个热点，今年大会的最佳论文就是探讨在线社区用户的生命周期和语言模式的变化。此外，还有很多使用社交网络数据进行研究的相关工作，如：社交网络数据搜索、社交网络中的信息推荐等。隐私保护问题也得到很多研究者的关注。

在网络搜索技术及相关用户行为分析领域，随着搜索交互技术的发展和用户行为分析技术的深入应用，搜索技术的研究重点已经由原有的以结果排序技术为单一重心，逐渐转变为在真实万维网搜索应用场景下对用户信息需求的分析与满足。以用户会话，而非单一查询为粒度进行算法和评价机制设计成为研究热点，而垂直搜索结果整合、多样化搜索等与用户实际需求密切相关、与传统结果排序技术迥然不同的技术研究也受到很多关注。同时，在社交网络、新闻搜索等上下文信息丰富的应用情境下如何设计搜索算法也有较多研究工作涉及。

主题报告

本届大会邀请了路易斯·梵阿纳 (Luis von Ahn)、米格尔·尼可雷里斯 (Miguel Nicolelis)、乔恩·克莱因伯格 (Jon Kleinberg) 3 位在产业界与学术界均有重

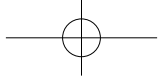
要影响力的专家作大会主题报告。

路易斯·梵阿纳博士是卡耐基梅隆大学计算机系副教授，创建了多个成功的创业公司。他是“人本计算” (human computation) 及众包技术的研究与实践先驱之一，获得过麦克阿瑟奖 (MacArthur Fellowship) 与史隆研究奖 (Sloan Fellowship)。在其主题报告中，梵阿纳博士介绍了其赢得广泛赞誉的 CAPTCHA 项目（通过为网站提供验证码服务的形式，借助用户行为实现 OCR 文本识别与纠错），并从借助用户力量服务反馈用户的理念出发，展示了正在开发运营的 Duolingo 项目。Duolingo 为广大学习外语的用户提供免费学习软件，并采用用户完成习题等形式调动用户群体力量对互联网外语内容进行翻译。

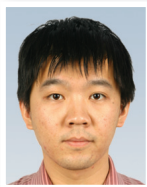
米格尔·尼可雷里斯博士是杜克大学医学院教授，在计算神经科学、尤其是脑机接口 (brain machine interfaces, BMI) 研究方面具有丰富的经验。在主题为“Computing with Brain Circuits”（基于脑芯片的计算）的报告中，他介绍了当前利用脑机接口技术实现大脑与机械、计算乃至虚拟设备交互的技术进展。通过介绍一系列的实验结果，他向与会人员展示了使用实时计算模型将人脑神经信号转化为机械运动指令的工作成果。

乔恩·克莱因伯格博士是康奈尔大学计算机系教授、美国科学院院士、著名的超链接引导的主题搜索算法 (hyperlink induced topic search, HITS) 的发明人。他在主题为“Computational Perspectives on Social Phenomena in On-Line Networks”（面向在线社会网络的计算模型）的报告中，重点介绍了社交网络中出现的群体行为和用户的网络交互。他着重从社会平衡理论、社会地位理论、结构多样性等方面介绍了在社交网络中发现的一些本质现象。用他本人的话来表述则是“the web knows you better than you know yourself”（网络比您自己更懂您）。此外，克莱因伯格博士还探讨了这些社会现象在社交网络分析中的应用。

最佳论文

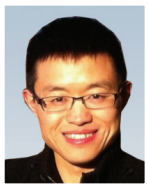


本年度的大会最佳论文是斯坦福大学的克里斯蒂安·德内斯库-尼库勒斯古-密吉尔 (Cristian Danescu-Niculescu-Mizil)、罗伯特·韦斯特 (Robert West)、丹·朱拉斯凯 (Dan Jurafsky)、朱雷·莱斯科夫 (Jure Leskovec) 和克里斯托弗·波茨 (Christopher Potts) 等人的 “No Country for Old Members: User Lifecycle and Linguistic Change in Online Communities”。该文从全新的视角研究了用户在从加入一个网络社区到离开该社区的过程中，自身的语言行为模式是如何变化的。他们发现大多数网络用户存在一个典型的语言接收、语言同步、到“语言青春期”（语言背离），直到最后离开该社交网站的过程。该文为解释社交网络中特定现象提供了理论和技术支持。



刘奕群

CCF会员。清华大学副教授。主要研究方向为网络信息检索、机器学习、用户行为分析等。yiqunliu@tsinghua.edu.cn



唐杰

CCF高级会员、本刊编委。清华大学副教授。主要研究方向为Web社会网络挖掘、语义Web等。jietang@mail.tsinghua.edu.cn

未来展望

随着网络应用研究的发展，近年来国际万维网大会所涉及的研究论文领域分布也发生了明显的变化。从近年来大会论文录用和参会的情况，可以看出几个有趣的趋势：万维网研究和应用更加“社会化”；万维网搜索和数据挖掘还会在相当一段时间继续保持稳定发展，同时有朝着个性化方向发展的趋势；另一个研究热点将是网络安全、隐私以及软件可靠性和可扩展性。

考虑到未来互联网环境和用户应用需求可能演变的形式，传统搜索引擎面对的数据对象和用户对象已经并将继续发生较大的变化。在数据层面，网络中蕴含的海量规模、高质量结构化数据对于搜索引擎成为不可或缺的重要信息来源，如何整合这部分信息对技术发展和产品运营提出了双重挑战；在用户层面，移动计算设备的普及与大数据革命带来的机遇，使搜索引擎对用户所处的上下文环境的认识也空前深刻，但如何将认识上的深入转化为用户信息获取过程的便利仍缺乏成功经验；除此之外，社交网络服务的兴起对互联网数据环境和用户群体将带来关键性影响，如何更好地面对相对封闭的社交网络数据环境和被社交关系组织起的用户群体，也是搜索引擎面临的机遇与挑战。■