

科研成果

一、学术专著

学术专著名称：《Moving Objects management: Models, Techniques and Applications》

学术专著作者：孟小峰 陈继东



该书是孟小峰教授及其团队历时十年的研究成果，全书比较系统地介绍移动对象管理相关内容，包括移动对象管理模型（包括移动对象建模、移动对象更新、移动对象索引等内容），移动对象管理技术（包括移动对象查询、移动对象预测、移动数据不确定性研究等内容），和移动对象管理应用（包括动态交通导航、动态交通网络、移动对象聚类分析、位置隐私保护等内容）等。

丹麦奥尔堡大学教授 Christian S. Jensen 为本书专门作序，给本书以高度的评价，认为 “The book meets the need for a coherent account of the state-of-the-art on important topics in the area of moving-object data management, which is at the core of the evolving mobile Internet. It comes highly recommended to research students and researchers new to the topics covered, as well as to experienced researchers”（该书全面系统地阐述了移动互联网的核心技术--移动对象数据管理的重要研究议题。值得高度推荐给研究生及相关研究人员）。Christian S. Jensen 教授目前担任国际数据库顶级期刊 VLDBJ 主编、国际数据库组织 SIGMOD 副主席，是移动数据管理领域著名专家。

二、论文集、专刊

论文集名称：Database Systems for Advanced Applications

论文集编辑者：Masatoshi Yoshikawa, Xiaofeng Meng 等

The 15th International Conference on Database Systems for Advanced Applications (DASFAA2010) 于 2010 年 4 月 1 日到 4 日在日本举行，会议论文集由 Springer 出版，孟小峰教授为本次会议的程序委员会成员，也是 DASFAA 2010 International Workshops: GDM, BenchmarX, MCIS, SNSMW, DIEW, UDM 论文集的编委之一。



论文集名称：Proceedings of the Second International Workshop on Cloud Data Management

论文集编辑者：Xiaofeng Meng, Ying Chen, Jiaheng Lu, Jianliang Xu

The Second International Workshop on Cloud Data Management (CloudDB 2010) 于 2010 年 10 月 30 日在加拿大多伦多举行，孟小峰教授担任本次研讨会的联合主席，并负责本次研讨会论文集的编辑工作。

专刊名称：Special Section on Trends Changing Data Management

专刊客座编辑：孟小峰教授，王海勋博士

计算机科学技术学报（JCST）Special Section on Trends Changing Data Management 专刊（2010 Vol.25 No.3）于 2010 年 5 月出版发行，孟小峰教授、王海勋博士受邀担任该专刊客座编辑。



三、论文列表

普适数据管理 (Pervasive Data Management)

- *C. Zhou, X. Meng: Out-of-Order Durable Event Processing in Integrated Wireless Networks. Accepted for publication in Journal of Pervasive and Mobile Computing (PMCIJ). (2010).
- C. Zhou, X. Meng: IO3: Interval-based Out-of-Order Event Processing in Pervasive Computing. In Proceedings of the 15th International Conference on Database Systems for Advanced Applications (DASFAA 2010): 261-268, April 1-4, 2010, Japan.
- 周春姐, 孟小峰, 文洁: Flickr 中的复合事件检测. 计算机研究与发展, 卷 47 (增刊): 1-7, 2010.10. (第二十七届中国数据库学术会议, 北京)
- 潘晓, 郝兴, 孟小峰: 基于位置服务中的连续查询隐私保护研究. 计算机研究与发展, 卷 47(1): 121-129, 2010.1.
- 周春姐, 孟小峰: 普适计算中复合事件检测的研究与挑战. 计算机科学与探索, 卷 4(12): 1057-1072, 2010.12.

云数据管理 (Cloud Data Management)

- *Y. Shi, X. Meng, J. Zhao, X. Hu, B. Liu, H. Wang: Benchmarking Cloud-based Data Management Systems. In proceedings of the CIKM Workshop on Cloud Data Management(CloudDB2010): 47-54, October 30, 2010, Toronto, Canada.
- J. Zhao, X. Hu, X. Meng: ESQP: An Efficient SQL Query Processing for Cloud Data Management. In proceedings of the CIKM Workshop on Cloud Data Management (CloudDB2010): 1-8, October 30, 2010, Toronto, Canada.
- X. Meng, Y. Chen, J. Lu, J. Xu: Report on the Second International Workshop on Cloud Data Management (CloudDB 2010). In Proceedings of the 19th ACM international conference on Information and knowledge management (CIKM2010): 1969-1970, October 30, 2010, Toronto, Canada.
- X. Meng, J. Lu, J. Qiu, Y. Chen, H. Wang: Report on the First International Workshop on Cloud Data Management (CloudDB2009).SIGMOD Record, Vol.39(1):58-60, March 2010.

闪存数据库系统 (Flash-based Database Systems)

- *X. Tang, X. Meng: ACR: an Adaptive Cost-Aware Buffer Replacement Algorithm for Flash Storage Devices. In Proceedings of the 11th International Conference on Mobile Data Management (MDM 2010): 33-42, May 23-26, 2010, Kansas City, Missouri, USA.
- D. Zhou, X. Meng: A Flash-Aware Random Write Optimized Database. In Proceedings of the 11th International Conference on Mobile Data Management (MDM 2010): 276-278, May 23-26, 2010, Kansas City, Missouri, USA.
- 汤显, 孟小峰: FClock: 一种面向 SSD 的自适应缓冲区管理算法. 计算机学报, 卷 33(8): 1460-1471, 2010.8. (第二十七届中国数据库学术会议, 北京)
- 卢泽萍, 孟小峰, 周大: HV-recovery: 一种闪存数据库的高效恢复方法. 计算机学报.(第二十七届中国数据库学术会议, 北京) (NDBC2010 “萨师煊优秀论文”)
- 梁智超, 周大, 孟小峰: Sub-Join: 面向闪存数据库的查询优化算法. 计算机科学与探索, 卷 4(5): 401-409, 2010.5.
- 周大, 梁智超, 孟小峰: HF-Tree: 一种闪存数据库的高更新性能索引结构. 计算机研究与发展, 卷 47(5): 832-840, 2010.5.

Web 数据管理 (Web Data Management)

- W. Liu, X. Meng, W. Meng: ViDE: A Vision-Based Approach for Deep Web Data Extraction. IEEE Transactions on Knowledge and Data Engineering(TKDE). Vol.22(3): 447-460 (2010).
- W. Liu, X. Meng, J. Yang, J. Xiao: Duplicate Identification in Deep Web Data Integration. In proceedings of the 11th International Conference on Web-Age Information Management (WAIM2010): 5-17, July 15-17, 2010, Jiuzhaigou, China.
- Y. Kou, Y. Li, X. Meng: DSI: A Method for Indexing Large Graphs Using Distance Set. In proceedings of the 11th International Conference on Web-Age Information Management (WAIM2010):297-308, July 15-17, 2010, Jiuzhaigou, China.
- W. Liu, X. Meng: A Holistic Solution for Duplicate Entity Identification in Deep Web Data Integration. In proceedings of the 6th International Conference on Semantics, Knowledge & Grids(SKG2010): 267-274, Nov. 1-3, 2010, Ningbo, China.

- Y. Li, D. Elswailer, X. Meng: Towards Task-Organised Desktop Collections. In Proceedings of the ACM SIGIR Workshop on Desktop Search: Understanding, Supporting, and Evaluating Personal Data Search (DS2010): 21-24 , July 23, 2010, Geneva, Switzerland.
- 黄静, 陆嘉恒, 孟小峰: 高效的 XML 关键字查询改写和结果生成技术. 计算机研究与发展, 卷 47(5): 841-848, 2010.5.

系统演示 (Demo)

- Y. Li, X. Zhang, X. Meng: Exploring desktop resources based on user activity analysis. In Proceedings of the 33rd Annual International ACM Special Interest Group on Information Retrieval Conference on Research and Development in Information Retrieval (SIGIR2010): 700 , July 19-23, 2010, Geneva, Switzerland.
- 胡享梅, 赵婧, 孟小峰, 王仲远, 史英杰, 刘兵兵, 王海平: TaijiDB: 一个双核云数据库管理系统. 计算机研究与发展, 卷 47 (增刊): 433-437, 2010.10.(第二十七届中国数据库学术会议, 北京) (NDBC2010 最佳系统演示)
- 黄毅, 潘晓, 孟小峰: OrientPrivacy: 移动环境下的隐私保护服务器. 计算机研究与发展, 卷 47 (增刊): 438-441, 2010.10.(第二十七届中国数据库学术会议, 北京)

注: 标'*'为年度代表论文。

毕业生学位论文

1. 李玉坤, 数据空间模型与查询技术研究 (Research on Data Space Model and Query Processing), 中国人民大学, 博士生毕业论文, 2010.5.6
2. 潘晓, 位置隐私保护技术研究 (Location Privacy Preserving), 中国人民大学, 博士生毕业论文, 2010.5.6
3. 周大, 闪存数据库系统存储和索引技术研究 (Storage and Indexing on Flash-based Database Systems), 中国人民大学, 博士生毕业论文, 2010.5.6
4. 王仲远, 面向领域的 Web 数据集成技术研究 (Research on Domain-Oriented Web Data Integration), 中国人民大学, 硕士生毕业论文, 2010.5.21
5. 寇玉波, 数据空间中图搜索技术的研究 (Research on Graph Searching Techniques in DataSpace), 中国人民大学, 硕士生毕业论文, 2010.5.21
6. 张相於, 数据空间索引关键技术研究 (Research on Dataspace Indexing Techniques), 中国人民大学, 硕士生毕业论文, 2010.5.21
7. 郝兴, 连续密度查询处理关键技术研究 (Research on Key Techniques of Continuous Density Queries), 中国人民大学, 硕士生毕业论文, 2010.5.21
8. 徐俊劲, 基于同义词规则的字符串近似搜索技术研究 (Research on Efficient String Similarity Search Using Synonyms), 中国人民大学, 硕士生毕业论文, 2010.5.21
9. 艾静, Web 信息可信性及用户隐私保护问题研究 (Research on Web Information Credibility and Privacy Preserving on the Searchable Internet), 中国人民大学, 硕士生毕业论文, 2010.5.21

四、专利

已授权专利：

1. 基于视觉的 Web 数据抽取系统和方法

发明名称：基于视觉的 Web 数据抽取系统和方法

申请人：孟小峰

专利号：ZL200810056103.4

获批时间：2010-2-17



2. 一种智能 Web 查询接口系统及其方法

发明名称：一种智能 Web 查询接口系统及其方法

申请人：孟小峰

专利号：ZL200810056104.9

获批时间：2010-08-13



已申请专利

1. 一种感知服务质量的位置隐私保护方法

发明名称：一种感知服务质量的位置隐私保护方法

申请人：孟小峰

申请号：201010193368.6

申请时间：2010-06-07



2. 一种防止位置依赖攻击的位置隐私保护方法

发明名称：一种防止位置依赖攻击的位置隐私保护方法

申请人：孟小峰

申请号：201010193366.7

申请时间：2010-06-07



3. 一种基于位置服务的连续查询隐私保护方法

发明名称：一种基于位置服务的连续查询隐私保护方法

申请人：孟小峰

申请号：201010195409.5

申请时间：2010-06-09



4. 一种基于闪存的自适应缓冲区置换方法

发明名称：一种基于闪存的自适应缓冲区置换方法

申请人：孟小峰

申请号：201010566968.2

申请时间：2010-11-25



5. 一种基于闪存的数据库恢复方法

发明名称：一种基于闪存的数据库恢复方法

申请人：孟小峰

申请号：201010552789.3

申请时间：2010-11-19



6. 一种基于短信终端的信息查询系统及查询方法

发明名称：一种基于短信终端的信息查询系统及查询方法

申请人：孟小峰

申请号：201010221373.3

申请时间：2010-6-29



五、科研项目

课题来源：国家自然科学基金面上项目

课题名称：Web 信息可信性研究

课题负责人：孟小峰

课题起止年限：2011 年 1 月 至 2013 年 12 月

课题简介：

随着网络与通信技术的迅速发展，Web 上的信息越来越多，并且仍以惊人的速度快速增长，Web 已经成为人们获取信息的重要途径之一。然而，这些信息纷繁复杂、鱼目混珠，如何鉴别这些信息的可信性已经成为一个日益重要的研究问题。本课题从基于内容分析的可信度评估机制、基于评分和投票的可信度评估机制、基于传播的可信度评估机制、基于数据源与记录关联分析的可信性评价机制四个方面深入分析了国内外信息可信度研究的现状。在此基础上，本课题拟从浅层网络和深层网络的角度出发，针对浅层网络和深层网络的特点和现有技术的缺陷，系统地研究 Web 上信息可信度的基本理论和实现方法，重点研究浅层网络中基于网页间关联的可信性研究、深层网络中基于数据源、数据记录以及双层关联机制的可信度研究四个方面的问题。通过本课题的研究为 Web 上信息可信度的进一步研究与应用提供理论方法、技术支撑和新的思路。

课题来源：国家“核高基”项目“非结构化数据管理”

课题名称：纯 XML 数据库系统

课题负责人：孟小峰

课题起止年限：2010 年 1 月 至 2011 年 12 月

课题简介：

主流非结构化数据类型包括音频、图像、视频、图形、文本等。这些数据都具有一定的语义特征或结构特征，而这些特征用 XML 来描述十分合适。非结构化数据管理系统方面，需要支持包括各种文档类型在内的媒体内容的全文检索和特征表达与查询。基于 XML 数据库实现内容的特征表达与查询，是一种可行的选择。此外，经过国内外研究人员的不断努力，到目前为止纯 XML 数据管理技术已相对成熟，利用 XML 数据库实现这些特征数据的管理成为可能。因此研究纯 XML 数据库系统分课题非常必要，它是顺利完成课题的基础，是课题不可缺少的一部分。分课题的课题目标包括以下 3 个方面：1）. 研发一套具有自主知识产权的软件，支持 XML 数据的多粒度存储，支持 XML 标准查询语言 XQuery 和 Xpath，提供 Java API。2）部署三项典型示范应用，将纯 XML 数据管理系统集成到非结构化数据管理系统中，并成功应用于新闻媒体，数字图书馆，流程工业三个行业示范应用 3）实现一套完整的基准测试平台，研究和建立一套完整的基准测试指标体系和测试平台。

课题来源： IBM 开放协作研究项目（Open Collaborative Research OCR）

课题名称： Cloud based Large Scale Data Management

课题负责人： 孟小峰

课题起止年限： 2010 年 至 2012 年

课题简介：

WAMDM 实验室团队在 2009 年 9 月，通过 IBM SUR 资助，开始承接了关于云数据管理系统的开发工作。这次由于团队的优秀表现，IBM 公司又提供了 IBM 开放协作研究（Open Collaborative Research OCR）项目资助，资助课题为“Cloud based Large Scale Data Management”。两项资助联合起来，共同搭建一个开源的云数据管理平台。

IBM 开放协作研究(Open Collaborative Research 简称 OCR)项目，是由 IBM 设立的一项科学研究资助计划，该计划旨在加强与国际一流高校科研人员的技术合作与交流，促进科研成果的快速转化。这次申请中，IBM 中国研究院向美国总部提交了十份申请。通过多轮讨论，答辩，表决，最后仅人民大学这个项目获得资助。这次我院的 OCR 项目在全球范围内申请成功，确实难度较大，成功来之不易。

相信通过这个项目的支持和团队成员的共同努力，必将使中国人民大学信息学院在国际云数据管理研究上占据一席之地，并带动国内相关研究工作的开展。

课题来源：MSRA Faculty Award

课题名称：Cloud-based Database System

课题负责人：孟小峰

课题起止年限：2010 年 至 2011 年

课题简介：

如何有效的管理大规模的数据在很多领域，例如：医疗保健，移动通信等，已经成为了一个很具有挑战性的问题。逐渐增大的数据量（甚至达到 PB 级别以上）对数据存储的体系结构，大规模的并行化查询处理，和数据分析处理提出了严峻的挑战。同时，大规模数据中心和计算机集群的兴起也产生了一个新的商业模式：云计算，基于云计算，公司和个人可以租借存储和计算能力，而不用花费大量的资本投资来构建和准备大规模的计算机安装。因此，基于云的数据存储和管理是一个快速发展的商业模式。在这个课题，我们设计基于云的数据库系统，用来支持下一代的信息管理和大规模分析处理的数据管理解决方法。课题的目标在于研究新的能够处理下一代的管理大规模数据的应用程序并且能够应用于多个领域（医疗保健，移动通信等）的数据库系统。

课题来源： 国家自然科学基金 青年项目

课题名称： 混沌人工神经网络的特性研究与混沌控制及其在联想记忆中的应用

课题负责人： 杨刚

课题起止年限： 2011 年 1 月 至 2013 年 12 月

课题简介：

混沌现象普遍存在于人脑活动中，混沌动态与人工神经网络的结合为实现真实世界智能计算提供了新契机，混沌神经网络在组合优化、联想记忆和人工智能等领域具有广阔的应用前景，各领域对混沌神经网络的适用性和混沌控制的灵活性提出了更高的要求。为了探索混沌神经网络动态特性，提高混沌神经网络性能，建立满足大规模联想记忆应用需求的自适应混沌神经网络模型，本课题拟基于数学推理和大量实验，利用数据统计分析的方法，形成混沌神经网络的特性分析结果，构造混沌神经网络的特性分布与参数调节模型，用于指导混沌神经网络的性能提高和混沌动态控制；构建灵活的自适应混沌动态控制策略，实现混沌神经网络中混沌动态的多运行轨迹变换，并将得到的特性模型和控制策略应用于联想记忆，提出具有自适应多对多联想记忆功能的混沌神经网络算法；同时研究增加网络存储容量和应用效率的优化算法，使得本课题的算法和模型适用于实际的应用系统。