

科研成果

一、学术专著

学术专著名称：《XML 数据管理：概念与技术》

学术专著作者：孟小峰

孟小峰教授所领导的课题组依据 10 年来在 XML 数据管理领域取得的研究成果，完成学术论著《XML 数据管理：概念与技术》。本书从数据库系统实现的角度，全面系统地介绍了纯 XML 数据库系统相关技术。内容涵盖了 XML 数据库存储管理技术（包括存储、编码、索引等方法）；XML 查询处理与优化技术（包括 XML 查询代数、结构查询处理、整体查询处理、查询优化等）；以及 XML 数据管理新技术（包括 XML/Update 处理、访问控制、关键字查询、近似查询处理等）；最后介绍了典型的 XML 数据库系统和基准测试。



学术专著名称：《移动数据管理：概念与技术》

学术专著作者：孟小峰 丁治明

孟小峰教授所领导的课题组依据多年的研究成果，完成学术论著《移动数据管理：概念与技术》。本书总结了国内外有关移动数据管理的研究工作和具有代表性的关键技术，详细介绍了课题组近年来的一些研究成果。内容涵盖嵌入式移动数据库技术（包括移动事务处理、移动数据库复制、移动数据广播等内容），移动对象管理技术（包括移动对象建模、移动对象索引、移动对象更新、移动对象查询、移动对象聚类、移动对象预测、移动数据不确定性研究等内容），位置相关的信息技术（包括位置相关数据的管理技术、空间数据库与交通网络数据库、位置隐私保护等内容）。

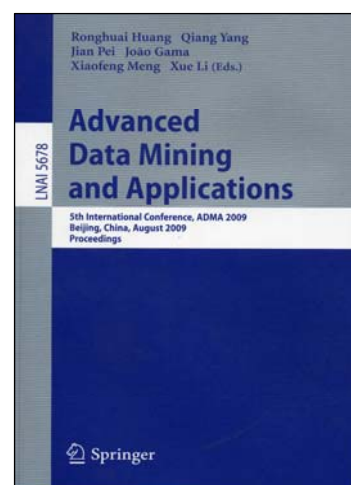


二、论文集

论文集名称：Advanced Data Mining and Applications

论文集编辑者：Xiaofeng Meng 等

第五届高级数据挖掘及其应用国际会议(ADMA 2009)于 2009 年 8 月 17 日-19 日在北京师范大学举行，该会议论文集由 Springer 出版，孟小峰教授为此次国际会议的联合主席，也是该会议论文集编委之一。



三、论文列表

1. X. Pan, X. Meng, J. Xu: Distortion-based Anonymity for Continuous Query in Location-Based Mobile Services. Accepted for publication in the proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL GIS 2009):256-265, November 4-6, 2009, Seattle, Washington.
2. W. Liu, X. Meng, and W. Meng: ViDE: A Vision-based Approach for Deep Web Data Extraction. Accepted for publication in IEEE Transactions on Knowledge and Data Engineering (TKDE).
3. J. Lu, J. Han, X. Meng: Efficient Algorithms for Approximate Member Extraction Using Signature-based Inverted Lists. In Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM2009):315-324, November 2-6, 2009, Hong Kong, China.
4. D. Zhou, X. Meng: RS-Wrapper: Random Write Optimization for Solid State Drive. In Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM2009):1457-1460, November 2-6, 2009, Hong Kong, China.
5. Y. Li, X. Meng: Supporting Context-based Query in Personal DataSpace. In Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM2009):1437-1440, November 2-6, 2009, Hong Kong, China.
6. X. Zhang, Z. Wang, J. Ai, J. Lu, X. Meng: An Efficient Multi-Dimensional Index for Cloud Data Management. In Proceedings of the CIKM Workshop on Cloud Data Management (CloudDB2009), November 2, 2009, Hong Kong, China.
7. Y. Li, X. Meng: Exploring Personal CoreSpace for DataSpace Management. In proceeding of 2009 Fifth International Conference on Semantics, Knowledge and Grid (SKG 2009):168-175, October 12-14, 2009, Zhuhai, China.
8. Y. Li, X. Meng, Y. Kou: An Efficient Method for Constructing Personal DataSpace. In Proceedings of the 6th Web Information Systems and Applications Conference (WISA2009), September 18-20, 2009, Xuzhou, China. (获优秀学生论文奖)
9. D. Zhou, X. Meng, Z. Liang: A New Cache Management Approach for Transaction Processing on Flash-based Database. In Proceedings of the 6th Web Information Systems and Applications Conference(WISA2009), September 18-20, 2009, Xuzhou, China.
10. Hui Ding, Nan Yang: An Approach of Community Detecting Based on Block Level Link Analysis. In Proceedings of the 6th Web Information Systems and Applications Conference(WISA2009), September 18-20, 2009, Xuzhou, China.

11. J. Zhou, X. Meng, T. Ling: Efficient Processing of Partially Specified Twig Pattern Queries, *Science in China Series F: Information Sciences*, Vol.52(10):1830-1847, Oct. 2009. Chinese version: 39(10):1034-1049.
12. J. Chen and X. Meng: Update-efficient Indexing of Moving Objects in Road Networks. *Geoinformatica*, Vol.13 (4):397-424, September, 2009.
13. J. Zhou, Z. Bao, T. Ling, X. Meng: MCN: A New Semantics towards Effective XML Keyword Search, In *Proceedings of the 14th International Conference on Database Systems for Advanced Applications (DASFAA 2009)*: 511-526, April 21-23, 2009, Brisbane, Australia.
14. F. Jiang, W. Meng, X. Meng: Selectivity Estimation for Exclusive Query Translation in Deep Web Data Integration, In *Proceedings of the 12th International Conference on Database Systems for Advanced Applications (DASFAA 2009)*: 595-600, April 21-23, 2009, Brisbane, Australia.
15. C. Zhou, X. Meng: Complex Event Detection in Pervasive Computing. The Third SIGMOD PhD Workshop on Innovative Database Research (IDAR2009), June 28, 2009, Providence, USA.
16. Z.Bao, T.W. Ling, B.Chen, J.Lu: Effective XML Keyword Search with Relevance Oriented Ranking. *IEEE International Conference on Data Engineering (ICDE2009)*: 517-528, March 29- April 3, 2009, Shanghai, China.
17. A. Behm, S. Ji, C. Li, J. Lu: Space-Constrained Gram-Based Indexing for Efficient Approximate String Search. *IEEE International Conference on Data Engineering (ICDE 2009)*:604-615, March, 29- April, 3, 2009, Shanghai, China.
18. J. Lu, Z. Bao, T. W. Ling, X. Meng: XML keyword query refinement. *Proceedings of the First International Workshop on Keyword Search on Structured Data.(KEYS 2009)*:41-42, June 28, 2009, Providence, Rhode Island, USA.
19. Shaoyi Yin, Philippe Pucheral, Xiaofeng Meng, PFilter: A Sequential Indexing Scheme for Flash-Based Embedded Systems, In *Proceedings of 12th International Conference on Extending Database Technology (EDBT2009)*: 588-599, March 23-26 2009, Saint-Petersburg, Russia.
20. 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. 第二十六届中国数据库学术会议论文集: 1-7, 2009.10. (第二十六届中国数据库学术会议, 南昌) (获萨师焯研究生优秀论文奖)
21. 王伟, 郭青松, 富丽贞, 孟小峰: 基于代数的 Transform 查询优化策略. *计算机研究与发展*, 第 46 卷(增刊): 74-80, 2009.10. (第二十六届中国数据库学术会议, 南昌)
22. 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. 第二十六届中国数据库学术会议论文集: 24-30, 2009.10. (第二十六届中国数据库学术会议, 南昌)

23. 艾静, 王仲远, 孟小峰: C-Rank: 一种 Deep Web 数据记录可信度评估方法. 计算机科学与探索, 第 3 卷, 第 6 期: 585-593, 2009.11.(第二十六届中国数据库学术会议, 南昌)
24. 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. 第二十六届中国数据库学术会议论文集: 68-74, 2009,10. (第二十六届中国数据库学术会议, 南昌)
25. 梁智超, 周大, 孟小峰: Sub-Join: 一种闪存数据库的查询优化算法. 第二十六届中国数据库学术会议论文集: 322-330, 2009,10. (第二十六届中国数据库学术会议, 南昌)
26. 寇玉波, 李玉坤, 孟小峰, 张相於, 赵婧: 个人数据空间管理中的任务挖掘策略. 计算机研究与发展, 第 46 卷(增刊): 446-452, 2009.10 (第二十六届中国数据库学术会议 南昌)

毕业生学位论文

1. 周军锋, XML 查询处理关键研究(Key Techniques on XML Query Processing), 中国人民大学, 博士生毕业论文, 2009.5.8。
2. 姜芳苕, Deep Web 数据集成中查询处理的若干关键问题研究(Key Techniques on Query Processing in Deep Web Data Integration), 中国人民大学, 博士生毕业论文, 2009.5.8。
3. 贾琳琳, 基于字典的近似查询匹配技术研究(Research on Fast Approximate Membership Checking), 中国人民大学, 硕士生毕业论文, 2009.5.12。
4. 黄静, XML 关键字查询技术研究(Research on XML Keyword Search Techniques), 中国人民大学, 硕士生毕业论文, 2009.5.12。
5. 朱金清, XML 近似查询处理技术研究(Research on XML Approximate Query Processing Techniques), 中国人民大学, 硕士生毕业论文, 2009.5.12。
6. 王伟, XML 查询关键技术研究(Research on Key Techniques of XML Query Processing), 中国人民大学, 硕士生毕业论文, 2009.5.12。
7. 向锂, 闪存数据库恢复技术研究(Research on Recovery Techniques for Flash-based DBMS), 中国人民大学, 硕士生毕业论文, 2009.5.12。

四、专利

已授权专利

1. 一种基于聚类来监控交通拥堵状况的系统及其方法

发明名称：一种基于聚类来监控交通拥堵状况的系统及其方法

申请人：孟小峰

专利号：2008100560991

申请时间：2008-1-11

获批时间：2009-9-25



2. 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：2008100561034

申请时间：2008-4-3

获批时间：2009-11-27



已申请专利

1. 一种个人数据空间环境下的任务挖掘系统和方法

发明名称：一种个人数据空间环境下的任务挖掘系统和方法

申请人：孟小峰

申请号：2009100900362

申请时间：2009-7-30



2. 一种基于用户特征的个人核心数据空间查询系统和方法

发明名称：一种基于用户特征的个人核心数据空间查询系统和方法

申请人：孟小峰

申请号：2009100900358

申请时间：2009-7-30



五、科研项目

课题来源：国家自然科学基金重点项目

课题名称：闪存数据库技术研究

课题负责人：孟小峰

课题起止年限：2009 年 1 月 至 2012 年 12 月

课题简介：

随着电子技术的发展，闪存作为新型存储介质被广泛应用于移动通信、工业控制、航空航天、笔记本电脑等嵌入式系统和便携式设备上。伴随着闪存容量的快速增长，闪存上的数据管理问题已成为新的挑战性问题。从数据库技术的角度研究闪存数据管理，构建新型的闪存数据库系统，对于促进闪存数据管理理论与闪存应用的进一步发展具有重要的理论意义和实用价值。本课题从闪存的器件特性入手，针对闪存应用的数据存取特点和现有数据库技术在闪存数据管理上的缺陷，从系统性和通用性角度研究闪存数据库的基本理论和设计方法，重点突破闪存数据库的体系结构、存储管理、查询处理、索引等关键问题，建立全新的闪存数据库系统理论和方法体系。课题从 DBMS 角度阐明闪存数据管理的特性，从高性能、高可用性、可剪裁性等方面构建闪存数据库的理论框架，为闪存数据库的进一步研究与应用提供理论方法和技术支撑，为数据库理论和技术的发展提供新思路。

课题来源：国家高技术研究发展计划(863 计划)重点项目

课题名称：普适计算基础软硬件关键技术及系统-隐私保护技术

课题负责人：杨楠 孟小峰

课题起止年限： 2009 年 3 月 至 2010 年 12 月

项目简介：

普适计算是一种超越桌面计算的新型计算模式，目标是在计算和通信无所不在的条件下建立以人为中心的计算环境——这种计算环境紧密结合于人们的生产生活环境，通过自然便捷的交互方式主动提供人们适用的信息服务，以辅助人们的活动。当前，信息技术日益普及，计算设施的泛在性逐渐显现，本项目的总体目标是研究分析制约普适计算技术和产业发展的关键技术问题，着重研发泛在设备自发互联互通的框架与协议，研发以人为中心的普适服务技术与系统，建立支持我国普适计算研究和相关产业发展的基础技术体系，项目研发产生的标准、协议和软硬件技术成果将促进相关产业的可持续发展。本项目组织产学研有机结合的研究团队，突破普适计算环境中设备、软件 and 用户三个层面上的基础性关键技术，构建支撑系统，并通过应用示范验证关键技术的有效性和促进产业进步。 WAMDM实验室负责项目中的异变环境中的数据管理技术和隐私保护匿名模型、算法和评价标准等关键技术。

课题来源：IBM 大学共享研究项目（IBM SUR 项目）

课题名称：Cloud-based Database Systems（云数据库系统）

课题负责人：孟小峰 陆嘉恒

课题起止年限：2009 年 9 月 至 2010 年 9 月

课题简介：

云计算是当今信息产业最受关注的一种计算模式，在这种模式下，企业和个人可以根据自己的需要购买存储设备和计算能力，而不是花费大量资金购买大规模高性能计算机。作为云计算的一项关键技术，云数据存储和云数据管理为业界带来巨大的潜在商用价值。随着信息产业的发展，企业和公司产生的数据量快速增长，通常数据规模可以达到TB甚至PB级别。如何管理和分析海量数据是目前很多领域所面临的问题，例如在医疗、通信和互联网领域。云环境是由大量的性能普通、价格便宜的计算节点组成的一种无共享大规模并行处理环境，所以从成本和性能两方面考虑，越来越多的企业更愿意把自己的数据中心从昂贵的高性能计算机转移到共有或私有云环境中。我们的项目主要致力于解决云计算环境下数据库的技术难点，实现一种具有高可用性、高容错性、可扩展性和高性能的云数据库系统。该项研究课题的目标是为了解决下一代海量分布式“大数据”（Big Data）的管理问题，从而支持下一代大数据应用系统以及大规模分析处理系统。在研究的第一年，将主要侧重于解决云计算平台上数据模型、存储与检索、同步与容错等一系列关键性问题。

课题来源：国家自然科学基金

课题名称：移动环境中关键词搜索的关键技术研究

课题负责人：陆嘉恒

课题起止年限：2010 年 1 月 至 2012 年 12 月

课题简介：

随着社会的不断发展，人们的移动性日益增强，对信息的需求也日益高涨，为有效地满足人们这种需求，在移动环境中搜索技术正逐步兴起。这种新兴的搜索是搜索技术在移动平台上的延伸，通过PDA、手机等移动通信终端，以短信、WAP 上网、语音通话等方式接入移动通信网络，来获取WEB、WAP站点信息、本地信息等信息服务。移动环境中搜索技术的出现，真正打破了地域、网络和硬件的局限性，满足了用户随时随地的搜索服务需求。而且搜索市场广阔，据统计，移动用户近是互联网用户的十倍。同时，移动增值服务业务的快速成长、互联网搜索技术的不断成熟以及3G 带来移动网络带宽为在移动环境中搜索行业的发展提供了机遇。因此，作为互联网搜索技术与移动通信技术相结合的产物，近年来，移动搜索技术在日益走向成熟，应用前景十分看好。本课题提出的“移动环境中关键词搜索的关键技术研究”，旨在研究面对移动环境中，关键词检索的一些关键技术，主要包括，(1) 定义高效的关键词检索语义。搜索结果简约，只有若干条，且高度符合用户需求；(2) 关键词的近似查询考虑同音替换和同义替换以包容数据输入中的错误；(3) 在移动环境下考虑位置的相关性，解决位置表示模型框架下的数据表示、存贮与索引，最近邻居查找的查询处理等关键技术问题。三者结合提供高效准确的移动搜索技术。我们的目标是基于可行的技术建立一个创新的面向移动环境中关键词检索原型系统。这个系统应该能够根据用户的位置，包容用户数据中的错误，实现近似查询。

课题来源：国家高技术研究发展计划(863计划)

课题名称：XML数据的复杂查询与模糊结构关键字检索

课题负责人：陆嘉恒

课题起止年限：2009 年 9 月 至 2010 年 12 月

课题简介：

可扩展标记语言（eXtensible Markup Language, XML）是 Internet 上一种新的数据表示、存储和交换标准。XML 推荐标准 1.0 版发布于 1998 年 2 月，之后迅速在全球掀起了 XML 应用的浪潮。XML 数据查询和检索在近十年以来，一直是数据库研究领域中的一个热点问题。虽然在国际和国内，企业界和学术界都投入了大量人力物力进行技术攻关，并开发出了一些原形系统，但是海量 XML 数据的高效复杂查询和检索问题仍远远没有得到完美解决，功能瓶颈和性能瓶颈两大问题仍大大制约着 XML 数据的广泛应用。所以我们相信对于 XML 数据的复杂查询和高效检索的研究，具有巨大的理论研究必要和现实应用价值。我们希望通过该课题的研究，在下一代 XML 数据管理技术方面有所突破，发表国际一流的学术论文，并结合我国的实际情况和产业需求，开发出具有自主知识产权的先进的 XML 查询和检索系统。不断提升我国数据库应用的水平，使之更好地为我国信息化建设服务。