

学术交流

一、孟小峰教授 2008 年学术活动任职

Program Committee Co-Chair, The 9th International Conference on Mobile Data Management ([MDM2008](#)), May 26-30, 2008, Beijing, China

Program Committee Co-Chair, IEEE International Workshop on Semantic Computing and Systems([WSCS2008](#)), July 14-15, 2008, Huangshan, China

Program Committee member, SIGMOD2008 The ACM SIGMOD International Conference on Management of Data ([SIGMOD2008](#)), June 9-12, 2008, Vancouver, Canada

Program Committee member, [DEXA2008](#) The 19th International Conference on Database and Expert Systems Applications(DEXA2007), 1-5 September 2008, Turin, Italy

Program Committee member, [ACM SAC2008](#) The 23rd Annual ACM Symposium on Applied Computing - Mobile Computing and Applications Track, March 16 - 20, 2008, Ceará, Brazil

Program Committee member, [PAKDD2008](#) The Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), May 20-23,2008, Osaka, Japan

Program Committee member, [APWEB2008](#) The 9th Asia-Pacific Web Conference, April 26-28, 2008, Shenyang, China

Program Committee member, [WAIM2008](#) The 9th International Conference On Web-Age Information Management, July 20-22, 2008, Zhangjiajie,China

Program Committee member, [WPS2008](#) The 2nd International Workshop on Web and Pervasive Security, March 17-21, 2008, Hong Kong

Program Committee member, [MobiDE 2008](#) The 7th ACM International Workshop on Data Engineering for Mobile and Wireless Access(MobiDE2008), June 13, 2008, Vancouver, British Columbia, Canada (in conjunction with SIGMOD 2008)

Program Committee member, [ODBIS2008](#) The 4th VLDB Workshop on Ontologies-based techniques for DataBases in Information Systems and Knowledge Systems(ODBIS2008), August 2008, Auckland, New Zealand

Program Committee member, [SCE 2008](#) The 11th IEEE International Conference on Computational Science and Engineering(IEEE SCE2008), July 16-18, 2008 - S?o Paulo – Brazil

Program Committee member, [ICWL2008](#) the 7th International Conference on Web-based Learning (ICWL 2008), 20-22 August 2008, Jinhua, China

Program Committee member, [WI2008](#) The 2008 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2008), 9-12 December 2008, Sydney, Australia

Program Committee member, [PALMS08](#) The 2nd International Workshop on Privacy-Aware Location-based Mobile Services (PALMS), In conjunction with the 9th International Conference on Mobile Data Management (MDM'08), April 27 2008, Beijing,

China

Program Committee member, [NIMC 2008](#) 2008 SIWN International Conference on Networking, Internet and Mobile Communications (NIMC 2008), 22-24 July 2008, Glasgow, UK

Program Committee member, [AMIGE2008](#) the IEEE Symposium on Advanced Management of Information for Globalized Enterprises (AMIGE2008), September 28-29, 2008, Tianjin, China

Program Committee member, [ICYCS 2008](#) the 9 th International Conference for Young Computer Scientists (ICYCS 2008), November 18-21, Zhang Jia Jie, China

二、孟小峰教授 2008 年学术交流

2008.06.07 ~ 2008.06.15 孟小峰教授和两位研究生李玉坤，郝兴参加在加拿大温哥华举办的国际数据库顶级会议 SIGMOD2008。



2008.10.25~2008.11.01 孟小峰教授和一位研究生潘晓参加在美国（Napa, Valley, California）举办的国际数据库会议 CIKM2008。

2008.10.24~2008.10.25 孟小峰教授和实验室多位同学参加了在桂林举办的第 25 届全国数据库学术会议。（NDBC2008）。





2008.11.08 孟小峰教授与多名研究生参加在安徽举办的 2008workshop on Flash-based Database.



三、2008 年专家来访

2008.04.22 周小冬博士 (AOL China) 应孟小峰教授邀请到实验室做题为 **Discovering and Summarizing Email Conversations** 的报告。

2008.06.23 实验室邀请李成锴博士 (University of Texas at Arlington) 进行学术报告: **Search the Database and Query the Web: Two Sides to the Story**。

2008.07.14 Qiong Luo 博士 (Hong Kong University of Science and Technology) 应孟小峰教授邀请到实验室做学术报告: **Relational Joins on Graphics Processors**。

2008.7 月初 美国纽约州立大学 Binghamton 分校孟卫一教授访问 WAMDM 实验室

2008.7 月初 美国 IBM T. J. Watson Research Center 的研究员 Haixun Wang 访问 WAMDM 实验室

2008.8 月初 加拿大 Simon Fraser University 的 jianPei 副教授访问 WAMDM 实验室

四、实验室研究生对外交流

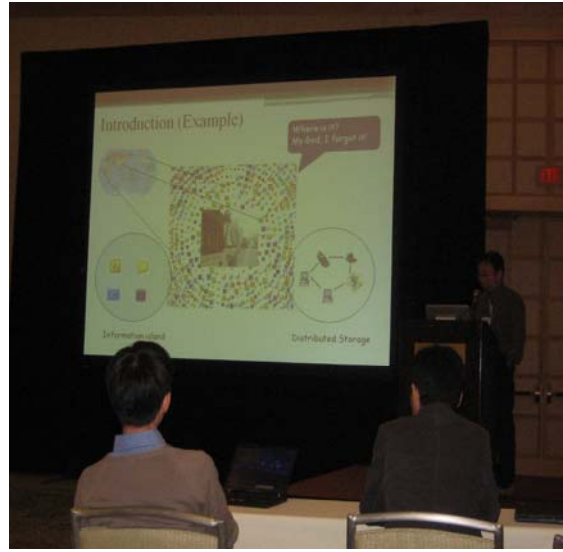
2008.05-2008.08 博士研究生周军峰到新加坡国立大学进行访问学习。



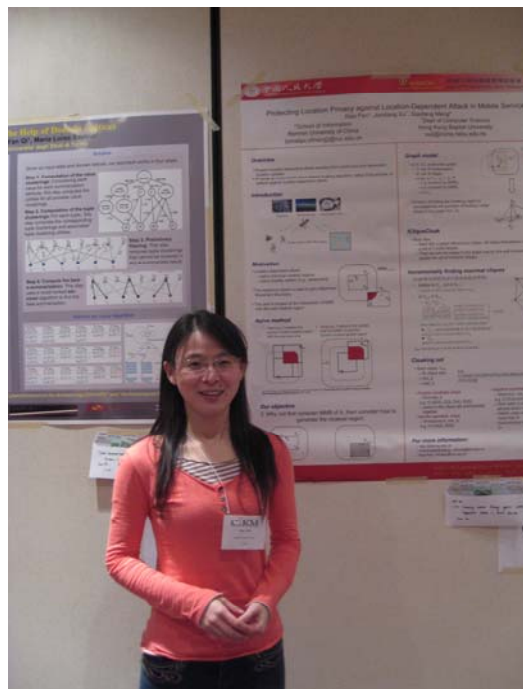
2008.06-2008.07 博士研究生潘晓赴香港浸会大学访问学习。

2008.10.24-2008.10.25 实验室多位同学参加了在桂林举办的第 25 届全国数据库学术会议 (NDBC2008)。

2008.6.7-2008.6.15 博士研究生李玉坤、硕士研究生郝兴参加在加拿大温哥华举办的国际数据库顶级会议 SIGMOD2008。



2008.10.25-2008.11.01 博士研究生潘晓参加了于美国 (Napa,Valley,California) 举办的国际数据库会议 CIKM2008。



五、承办会议

2008.04.27-2008.04.30

由 WAMDM 实验室主办的“第九届移动数据管理国际会议 (The 9th International Conference on Mobile Data management, 简称 MDM2008), 在中国人民大学逸夫会议中心举行。



移动数据管理国际会议是数据库领域高水平的专业会议，创始于 1999 年的香港，之后每年举行一次。汇集了国际著名大学、主流研究机构、跨国企业一流学者和产业界精英，引领着该领域技术发展的潮流。先后在香港、新加坡、澳大利亚、美国、日本、德国等地举办，这是该会议首次来到中国内地举办。

MDM2008 汇集了国际著名大学、主流研究机构、跨国企业一流学者和产业界精英，引领着该领域技术发展的潮流。汇集了全球知名专家，香港科技大学、本会议的创始人 Dik Lee 教授，澳大利亚 Monash 大学、本会议的指导委员会主席 Arkady Zaslavsky 教授，以及来自美国、德国、法国、日本、韩国、新加坡、瑞典、芬兰、挪威、奥地利、希腊、印度、西班牙、土耳其、丹麦、塞浦路斯、英国等 20 个国家和地区的一百多位海外学者和来自北京大学、中科院软件所、浙江大学、中国科大、东北大学等十多位国内学者参加本次了会议。

本次会议提供了向海外学者宣传和展示人民大学及信息学院的极好的机会，也是今年信息学院成立 30 周年重要活动之一。冯惠玲副校长代表学校到会致辞，向与会者介绍了人民大学学科特色和近年的发展情况。



孟小峰教授应邀担任本次会议的程序委员会主席，在会议上代表程序委员会介绍论文评审和会议组织情况。这是中国学者首次担任该主席。



孟小峰教授多年在移动数据管理方面做出突出成绩，1999年率先在国内开展移动数据库的研究工作，获得国内外学术界的认可，他同时是该会议的指导委员会委员。

作为高水平的学术会议，MDM2008共收到来自世界22个国家和地区的论文119篇，录用高水平学术论文26篇，论文录用率21.8%。会议设置了8个分组报告（research sessions），4个技术研讨会（workshops），2个辅导报告（seminars）。会议邀请了法国INRIA的Denis Caromel博士，美国Bell实验室Richard Hull博士，中国Google的Hanping Feng博士做特邀报告。会议内容涵盖了移动数据管理的若干方面，如移动对象管理，移动用户查询，情境感知，隐私与安全，传感器网络，数据广播等。

MDM2008 详细信息请访问大会的主题网站（<http://idke.ruc.edu.cn/mdm2008/>）

六、实验室研讨会

WAMDM 实验室每周六举办一次研讨会，研讨会内容包括研究工作进展报告、新技术介绍、参会情况汇报、经验交流等。2008 年度共举办研讨会 24 次，共有 62 个报告，以下是本年度研讨会报告简介。

2008.01.05 Venue: FL2, Meeting Room, Information Building	
Xin Zhang (XML Group)	XQuery/Update processing Abstract: This talk focus on the pocessing technique of XQuery/Update whose draft is proposed by W3C in the past year. this talk try to process XQuery/Update based on XML algebra. And how to optimize Transform query based on XML algebra.
Yanyan Ling (Web Group)	Research of Entity Identification for Web data management Abstract: Entity identification means to find those records which refer to the same real world entity from two or more data sources. In the web data integration, the variety and variability of the data sources make the entity identification on public data a very challenge problem. Secondly, some special applications require implementing the entity identification on private data.
Min Xie (XML Group)	Integrity Auditing of Outsourced Data Abstract: Some research points on integrity auditing of outsourced data, thing needs to be done before graduation, and the near future work
Zhen Xiao (Mobile Group)	Privacy Protection in Location Based Service Abstract: This talk analyzes the privacy problem in location based service and the related works in state-of-art. We also propose two solutions for different privacy issues.
2008.03.01 Venue: FL2, Meeting Room, Information Building	
Min Xie (XML Group)	LaTeX: An Introduction Abstract: A brief introduction of LaTeX.
Yukun Li (Web Group)	An efficient method for personal dataspace integration Abstract: An adaptive strategy for personal data integration is proposed, and a prototype system is developed and demonstrated in the paper.
Zhongyuan Wang (Web Group)	A New Method for Deep Web Data Integration Abstract:

	We propose a new method for Deep Web Data Integration which can implement precise extraction
2008.03.08 Venue: FL2, Meeting Room, Information Building	
Jinqing Zhu (XML Group)	Relaxed Twig Query Processing Abstract: Twig query processing can achieve high performance when compared to binary join processing. But two problems exists.
Junjin Xu (XML Group)	A New Query Semantic Abstract: This presentation introduces several query semantic of keyword search on XML document and their disadvantage , and we also define a new query semantic.
Linlin Jia (Web Group)	Automatic Scientific Paper Classification Abstract: Automatic scientific paper classification has become an important research topic due to the increasing number of large scientific paper collection. Several works have explored this problem but still have some drawbacks and the results can not meet the users's need in practice. We focus on improving performance of current works by search engin.
2008.03.22 Venue: FL2, Meeting Room, Information Building	
Xiangyu Zhang (Web Group)	An Event Based Email Processing Approach Abstract: Introduced an approach to help users process their emails by showing them collection of events rather than plain text.
Junfeng Zhou (XML Group)	Discussion of recent work of XML group Abstract: In this talk, I present the ongoing work of XML group and propose some interesting problems encountered in practice to discuss with all members of our lab.
2008.03.29 Venue: FL2, Meeting Room, Information Building	
Zhongyuan Wang (Web Group)	Introduction to Cloud Computing Abstract: This presentation introduced Cloud Computing which is looking like a classic disruptive technology. We also show the relationship of Web2.0, Grid Computing and Cloud Computing. At last, we discuss several Cloud Computing cases and the future of Cloud Computing.
Li Xiang (Mobile Group)	FlashDB & LazyHash Abstract: One problem of designing index for flash-based storage is that the hardware platform and workloads are quite very different. FlashDB and

	LazyHash try are optimized for all situations by self-tuning theirself.
2008.04.16 Venue: FL1, Meeting Room, Information Building	
Yubo Kou (Web Group)	An Introduction to Desktop Search Abstract: This presentation reveals a general perspective of current desktop search tools and introduces current research in this field.As this field becomes hotter and its significant relationship with PIM, desktop search deserves our attention.
Jing Huang (XML Group)	Dagstuhl Seminar on Ranked XML Querying Abstract: This report introduces some interesting talks about ranked XML querying presented in Dagstuhl Seminar.
2008.04.23 Venue: FL1, Meeting Room, Information Building	
Junjin Xu (XML Group)	Keyword Search on XML Tree Abstract: This presentation introduces Keyword Search on XML Tree, including the construction of the index, the processing of the query and the analysis of the result.
Da Zhou (Mobile Group)	A new IO mechanism for transaction processing on flash database Abstract: With the development of widely used PDA, MP3 and DC, flash memory, as a new electronic storage device, becomes more and more popular and important. There is an increasing trend of using database to manage the increasing data on flash memory. Moreover, transaction processing is needed to assure the correctness of complicated applications. Based on the analysis of traditional transaction IO mechanism, this paper presents a new IO mechanism of transaction.
2008.05.31 Venue: FL1, Meeting Room, Information Building	
Yukun Li (Web Group)	Research on Personal Dataspace Management Abstract: Explosion of the amount of digital information has made Personal Information Management(PIM) become a hot topic. Personal data is always distributed, rough-and-tumble, personalized, heterogenous and evolutionary, which brings much challenge to effective and efficient Personal Dataspace Management (PDSM). In the paper, by highlighting the importance of users in Personal Dataspace Management System(PDSMS), we proposed a user-centered framework. We first show research issues, related work, main research problems and challenges in this area. We then introduce the

	current research work and the preliminary results. Finally, the research plan of my PhD project is presented for discussion.
Xing Hao (Mobile Group)	<p>Continuous Density Queries for Moving Objects</p> <p>Abstract: A density query returns the regions with a density higher than some user-specified threshold. Although many studies have been done on density queries for moving objects in highly dynamic scenarios, they all focused on how to answer snapshot density queries for moving objects. This presentation proposed an approach which continuously monitors dense regions for moving objects.</p>
2008.07.05 Venue: FL1, Meeting Room, Information Building	
Jiaheng Lu (XML Group)	<p>Efficient Merging and Filtering Algorithms for Approximate String Searches</p> <p>Abstract: I present my research experience and present this ICDE paper, where we propose several new algorithms for efficient merging and filtering algorithms for Approximate String Searches.</p>
Linlin Jia (Web Group)	<p>Introduction to online advertising and AdCenter labs</p> <p>Abstract: A brief introduction to online Advertising and the demo developed by Microsoft AdCenter labs.</p>
Linlin Jia (Web Group)	<p>Faceted Search</p> <p>Abstract: A brief introduction to faceted search and the differences between faceted search and others.</p>
Kouyubo (Web Group)	<p>Introduction to Freebase</p> <p>Abstract: Freebase claims to be an open, shared database of the world's knowledge. Compared with Wikipedia, the information it stores and organizes is more structural. Freebase is a novel application of WEB2.0 and semantic web.</p>
2008.07.19 Venue: FL1, Meeting Room, Information Building	
Yukun Li (Web Group)	<p>EASE: An Effective 3-in-1 Keyword Search Method for Unstructured, Semi-structured and Structured Data</p> <p>Abstract: This is a paper published in SIGMOD2008(Guoliang Li, etc). In this paper, an efficient and adaptive keyword search method is proposed, called EASE, for indexing and querying large collections of heterogeneous data. It proposes an extended inverted index to facilitate keyword-based search, and presents a novel ranking mechanism for enhancing search effectiveness.</p>

Jinqing Zhu (XML Group)	Keyword Proximity Search in Complex Data Graphs Abstract: This is a sigmod'08 paper about keyword proximity search on graphs. Previous approach try to answer queries on graph by way of solving Steiner-tree problem. But they have two major flaws.
Zhiyong Xi (XML Group)	Template of the Report of PG Code Review Abstract: Template of the Report of PG Code Review.
2008.07.26 Venue: FL1, Meeting Room, Information Building	
Dingjie Chen, Wei .Chen	C-DBLP Project Presentation Abstract: C-DBLP Project is a Chinese computer science document-integration system based on data integration technology.The system is developed by Web Group in the Lab of Web and Mobile Data Management(WAMDM), Renmin University of China. Data in C-DBLP system which include papers published in famous journals and conferences are author-oriented organized to provide efficient and easy-to-use document retrieval services.
Xiangyu Zhang (Web Group)	Introduction to Progress of OrientSpace Abstract: Introduced some development progress of OrientSpace in this summer.
Zhiyong Xi (Undergraduate)	An example of the report of PG code review--executor Abstract: An example of the report of PG code review—executor.
2008.09.06 Venue: FL1, Meeting Room, Information Building	
Xiao Pan (Mobile Group)	Summary for research in Hongkong Abstract: This presentation is to share my work and Dr.Xu's work in HongKong with our lab.
Junfeng Zhou (XML Group)	Summary of Recent Work Abstract: In this talk, I mainly discuss recent work done in NUS, which consists of two aspects, the first is about extending XPath and was submitted to ICDE09, the second is about XML keyword search and will be submitted to DASFAA.
2008.09.13 Venue: FL1, Meeting Room, Information Building	
Chunjie Zhou (Mobile Group)	A Summary of Indoor Navigation Abstract: The trandational locating methods cannot be used in the indoor environment because of the influence of the signal strength, the accuracy, and so on. There are a lot of difference between indoor

	navigation and outdoor navigation, so the research of indoor navigation is becoming a hot point. This paper analyze the challenge problems of the indoor navigation. It summarizes the current researches in the aspect of navigation patterns, navigation techniques and navigation systems. Finally it points out the future research directions.
Xian Tang (mobile Group)	Summary of Recent Work Abstract: In this talk, I mainly introduce the simple idea of recent work. The idea comes from papers of two different aspects. One is about self-tuning and the other is about storage strategy.
2008.09.20 Venue: FL1, Meeting Room, Information Building	
Linlin Jia (Web Group)	Approximate membership lookup Abstract: How to efficiently extract all substrings from input documents such that the substring approximately matches a record in dictionary.
Yukun Li (Web Group)	OwnerCorrelation: A new Framework for Personal Dataspace Management Abstract: Correlation to owner is the root character of each object of PDS, and may play an important role in data operation of PDS. Based on the assumption, we propose a new concept OwnerCorrelation(OC) to describe the relation between owner and other objects of PDS, and present to take Personal Taskspace(PTS) to model characters of owner entity of PDS, which provides people a new position to view techniques of PDSMS.
2008.10.11 Venue: FL1, Meeting Room, Information Building	
Jinqing Zhu (XML Group)	Efficient Processing of XML Twig Query based on Related Semantics(For NDBC2008) Abstract: Though keyword search method can be used easily, it possesses the inherit feature of limited expressive capability. Structured query language has the powerful expressive ability, however, users must have a full understanding of the underneath schema information. We propose an extension to XPath by introducing a novel related semantics and an efficient algorithm rTwigStack based on this semantics.
Zeping Lu (Mobile Group)	The TPC-C and Current Testing Benchmark Abstract: TPC Benchmark TM C(TPC-C)is an OLTP workload, which is considered as performance evaluating standard. To evaluate software's performance, a TPC-C testing system is needed.
2008.10.17 Venue: FL1, Meeting Room, Information Building	

Xiangyu Zhang (Web Group)	Progress on Dataspace Abstract: Introduced our progress on system development and research on Dataspace.
Xiangyu Zhang (Web Group)	TEXEM: An Entity-based Task Extraction Approach for Emails(For NDBC2008) Abstract: An introduction to our work on task extraction from emails that is to be presented in NDBC 2008.
Zhongyuan Wang (Web Group)	A Data Driven Approach for Automatic Wrapper Generation and Maintenance(For NDBC2008) Abstract: This paper proposes a novel method to perform this issue
2008.10.31 Venue: FL1, Meeting Room, Information Building	
Kou Yubo (Web Group)	Experimental Results On Approximate Membership Checking Abstract: This experiment is about the problem of identifying sub-strings of input text strings that approximately match with some member of a potentially large dictionary. This problem arises in several important applications such as extracting named entities from text documents and identifying biological concepts from biomedical literature.
Jinqing Zhu (XML Group)	Analysis of SQL/XML Abstract: This PPT mainly concerned with the character of SQL/XML, especially the position of XML query in SQL. Then an initial solution is given to integrate the XML query into existing relational DB.
Yulei Fan (Mobile Group)	The Declarative Programming Language: Ruby Abstract: Ruby is a language of careful balance. Its creator, Yukihiro "matz" Matsumoto, blended parts of his favorite languages(Perl, Smalltalk, Eiffel, Ada, and Lisp) to form a new language that balanced functional programming with imperative programming. Important Language Features include Iteration, Expression, String, Controlling Structure, everything is an object, naming conventions, method access, Class, Singleton methods, Block, Exception Processing and Thread, and so on.
2008.11.07 Venue: FL1, Meeting Room, Information Building	
Xiao Pan (Mobile Group)	Summary for Attending CIKM08 Abstract: This presentation is to share our trip experiments with our lab. It contains number of papers&accept rate, three keynotes, some

	interesting talks and poster.
Junfeng Zhou (XML Group)	Revisiting XML Keyword Search Abstract: In this talk, I mainly discussed existing keyword search methods, then I introduced the initial idea of keyword search in XML stream.
2008.11.14 Venue: FL1, Meeting Room, Information Building	
Da Zhou (Mobile Group)	2008 workshop on flash-based database Abstract: 2008 workshop on flash-based database was hold in hefei, Anhui. More than twenty researchers took part in this workshop. The topics cover storage, index, query processing and transaction processing. We show our recent research about query on flash disk after the introduction of workshop. Some naive ideas are introduced and initial experiments are done. Finally we show some pictures in Hefei.
Qingsong Guo (XML Group)	Conditional Random Fields Model Abstract: A conditional random field (CRF) is a type of discriminative probabilistic model most often used for the labeling or parsing of sequential data, such as natural language text or biological sequences. And I'm trying to use it into XML keywords refinement.
Bingbing Liu (Web Group)	Cumulated gain-based evaluation of IR techniques Abstract: The author develops 4 new measures to evaluate the efficiency of different IR techniques which are CG, DCG, NCG, NDCG. The first one accumulates the relevance scores of retrieved documents along the ranked result list. The second one is similar but applies a discount factor to the relevance scores in order to devaluate late-retrieved documents. The third (fourth) one computes the relative-to-the ideal performance of IR techniques, based on the (discounted) cumulated gain they are able to yield. Then the author examines five different IR techniques using the new measures based on data of TREC-7 and discusses the parameters and the limitations of the new measures.
2008.11.29 Venue: FL1, Meeting Room, Information Building	
Xiangyu Zhang (Web Group)	2nd Stage Developpe Plan of OrientSpace Abstract: Introduced the developpe plan of OrientSpace in the second stage.
Xiangyu Zhang (Web Group)	Evolution of Personal Dataspace using User Feedback Abstract: Evolution is one of the most important features of dataspace systems. Evolution in personal dataspace is different from that of dataspace in data integration context. We propped a evolution framework for personal dataspace using user feedback. This framework can perform

	the evolution in a pay-as-you-go fashion.
Yukun Li (Web Group)	EASE: TaskSpace: A Task Based Model For Personal Dataspace Management Abstract: In this paper, we propose a task based model for organizing personal data items.
2008.12.07 Venue: FL1, Meeting Room, Information Building	
Wei Wang (XML Group)	Introduction to OrientX Abstract: An brief introduction to OrientX database system, including the architecture, main features, storage management, demonstration of OrientX3.0 and etc.
Jinqing Zhu (XML Group)	OrientX3.0 and its improvements Abstract: This ppt mainly deals with the XQuery implementation in OrientX3.0, including navigation-based query processing and algebra-based query processing. Then, the current efficiency problem in current version is analyzed. Finally, we show the implementation issue of XQuery/Update.
Wei Wang (XML Group)	Features of new version and plans Abstract: The feature definitions of OrientX new version and its plans.
2008.12.13 Venue: FL1, Meeting Room, Information Building	
Xiaowei Gu (Mobile Group)	Uncertainty Reasoning Under Correlative Knowledge Abstract: Uncertainty Reasoning infiltrate into many aspects of human life. First we considered the key components of a uncertainty reasoning system, then proposed a reasoning under correlative knowledge framework.
Chunjie Zhou (Mobile Group)	The probabilistic complex event detection in pervasice computing Abstract: Extracting complex events from low-level atomic events is becoming more and more important in daily life. However, current event detection researches often assume the data is precise. In many of the real applications, the data is instead imprecise. In addition, few research works have been carried out on finding better usages of time information. Due to the importance of event queries and the rapidly increasing amount of probabilistic data collected, we propose some temporal semantics, data model and event query techniques.
2008.12.20 Venue: FL1, Meeting Room, Information Building	
Jing Ai	Information Credibility on the Web

(Web Group)	<p>Abstract:</p> <p>Credibility on the Web is an important research topic. This presentation is a survey to introduce the concept of information credibility. Methods of identifying credibility in six different Web scenarios are also introduced detailed, including P2P network, online discussion forums, Wikipedia and so on. Besides, information credibility criteria and two typical evaluation methods are introduced. At last, a brief summary about information credibility on the Web.</p>
Xiangmei Hu (Web Group)	<p>Web Spam</p> <p>Abstract:</p> <p>A survey is done on currently popular web spams, which introduces two spam techniques</p>
<p>2008.12.27 Invited Talks Venue:FL1,Meeting Room,Information Building</p>	
Jidong Chen (EMC Research China)	<p>Cloud based Personal Information Management – Introduction to EMC Cloud Computing</p> <p>Abstract:</p> <p>With the requirements of automatic online storage and backup, round the clock access and securely sharing and publishing of personal digital information, it is inevitable that personal information management will migrate into the cloud. The goal of personal information cloud service is to securely access and organize all your information anytime, anywhere, using any device and never lose any of it. EMC is creating a new cloud services business called Decho ('digital echo' referring to the reverberating accesses to information in a user's digital environment) by joining Mozy (cloud backup) and Pi (personal information) together. It will use EMC data centres around the planet to store consumer and business files using Mozy's software front end to provide data ingest and access services and Pi's metadata software to manage and verify personal information. Decho can deliver on the promise of cloud-based personal information management and can help individuals everywhere preserve, manage and enrich the information most important to them.</p>
Yunpeng Chai (Tsinghua University)	<p>Understanding and Comments on Cloud Computing</p> <p>Abstract:</p> <p>Cloud computing is new concept proposed in recent years. This talk firstly compares cloud computing with traditional distributed computing and grid computing to help understanding the concept and characteristics of cloud computing, and then introduces some possible research directions in both cloud computing platform and combining with applications.</p>

七、企业合作

合作单位：EMC 中国实验室

合作题目：*Task Management: Exploring Personal Information Management of Email in Mobile Environment*

合作方介绍：EMC 公司是全球唯一专注于提供智能化企业级存储解决方案的公司。总部在美国马萨诸塞州的 EMC 公司是信息存储、管理、软件、服务和解决方案领域世界领先的公司,在世界 500 强企业中排 262 位，全球 10 大软件公司第 6 名。

合作内容介绍：我们和 EMC 的合作关注于个人数据空间的研究，具体来说，希望利用数据空间技术解决个人数据管理目前面临的问题。研究课题涉及个人邮件管理相关技术、个人数据空间理论及模型、原型系统开发等多各方面。合作开展以来，于今年 1 月和 10 月举行了两次研讨会，目前在理论研究和系统开发方面，取得了一些研究成果。

研讨会简介：

1. 2008 年 1 月 25 日召开了第一次研讨会。参加人员包括 EMC 中国实验室首席科学家毛文波博士，EMC 中国实验室研究员陈继东博士，人民大学孟小峰教授，以及其他项目组成员。会议主要讨论了个人邮件管理相关技术，认为作为个人数据空间的重要数据源，针对邮件管理技术的研究具有重要意义。研讨会上 EMC 研究人员介绍了公司在个人数据信息管理方面的研究规划。我们报告了关于邮件中任务提取策略方面的工作，提出充分利用邮件中的结构和实体信息，将电子邮件通过实体聚类的方法转化为任务的集合，并对任务的重要性进行评估，从而将无结构的邮件内容转化为具有结构的任务集合，方便了用户处理邮件的过程。此外还介绍了我们在个人数据空间管理方面的设想和初步成果。
2. 2008 年 10 月 25 日，双方举办了第二次研讨会。参加人员包括：EMC 公司信息安全部首席科学家 Burt Kaliski 博士，EMC 中国实验室首席科学家毛文波博士，EMC 中国实验室研究员陈继东博士，孟小峰教授以及项目组其他成员。我们介绍了在个人数据空间方面取得的一些研究成果和准备在数据空间集成、数据空间模型方面开展的研究工作。介绍了我们提出的核心数据空间模型 Corespace，并演示了我们开发的原型系统 OrientSpace，对系统中的特色功能进行了讨论和交流。介绍了我们关于数据空间演化集成（pay-as-you-go）的研究工作。讨论认为这是一个数据空间研究中十分重要的问题。我们的工作得到了对方的积极评价，双方讨论了下一步工作。

合作单位：华为技术有限公司

合作题目：基于 Flash 的数据管理系统研究

合作方介绍：华为是全球领先的电信解决方案供应商，专注于与运营商建立长期合作伙伴关系，其产品和解决方案涵盖移动、核心网、网络、电信增值业务和终端等领域。随着网络和信息技术的发展，华为的业务范围不断拓展，已覆盖到数据存储、网络安全以及综合业务软件等技术领域。

合作内容介绍：

1. 基于 Flash 的数数据管理评价指标。主要研究系统整体性能与 flash 读、写、擦等的次数之间有怎样的定量关系；系统电源消耗与读、写、擦次数之间的定量关系；内存占用和性能优化二者之间的关系。在任何给定大小的内存限制下，

系统都能够最大化利用资源从而使得性能最优。也就是说，系统实现不依赖于内存大小，同时又能够最大化地利用已有内存。

2. 基于 Flash 的数据存储和索引模型（以及相应的垃圾回收方案）。研究在 flash 上如何组织记录和索引。具体来说，就是研究数据库一个表中原始记录和索引条目分别以何种方式存储在 flash 上；不同的表之间的数据存储位置关系；同一个表的不同索引结构之间的存储关系。研究在有数据库插入、删除和更新操作之后，怎样最小化 flash 的更新。研究缓冲策略，在尽量少使用内存的情况下最小化索引维护的代价。无论如何存储和维护，垃圾回收都是不可避免的，所以在不破坏损耗平衡的同时，研究垃圾回收策略来顺应系统的总体目标。
3. 基于 Flash 的数据的数据库系统框架。在上述研究基础上，加入数据库的查询处理，事务处理，安全访问控制和数据加密算法，并考虑支持并发操作。确定适合 Flash 特性的数据库系统框架。