附 录

实验室研讨会

2010.12.10 Venue: FL1, Meeting Room, Information Building	
Zhichao Liang (Flash Group)	A novel method to extend flash memory lifetime in flash-based DBMS Abstract: As the capacity increases and the price drops gradually, flash memory is becoming the promising replacement of disk, even in the enterprise applications. However, flash memory suffers from erase-before-write and limited write-erase cycles at the same time, which means the abuse of write,especially small and random write, will wear a flash block out quickly. We analyze the free space management in traditional DBMS and point out its disadvantage when used on flash device. In addition, we also propose a new solution involving free space management and buffer management to extend the lifetime of flash memory by reducing the number of write I/O.
Xiaoying Qi (Flash Group)	An Operation Aware Flash Translation Layer for Enterprise-class SSDs Abstract: Flash translation layer is an important firmware in flash-based devices. It is critical to affect the performance of flash-based devices. So when SSDs are used in enterprise-class environment, FTL should be redesigned to improve the whole performance. In this report, we introduce an operation aware flash translation layer for enterprise-class SSDs.
2010.12.03 Venue: FL1, Meeting Room, Information Building	
Wei Tong (Web Group)	A Structured Approach to Query Recommendation With Social Annotation Data [ppt] Abstract: Query recommendation has been recognized as an important mean to help users search and also improve the usability of search engines.
Sen Yang (Web Group)	Introduction to OpenScholar Abstract: OpenScholar is a web system to build scholars' homepage automatic. Its features of searching scholars' infomation and dynamic maintenance can help users build their homepages easily and fast.
2010.11.26 Venue: FL1, Meeting Room, Information Building	

Haiping Wang (Cloud Group)	**Research of query optimization in the cloud** Abstract: In cloud data management systems,data is partitioned into blocks and replicated.It is nesscary to translate some data blocks when we do some types of query processing.So we did some research on how to finish the query with little costs.
Xiaojian Zhang (Web Group)	**Record Linkage with Uniqueness Constraints and Erroneous Values** [ppt] Abstract: This paper presents some challenges of record linkage and data fusion in heterogeneous data sources with uniqueness constraints and erroneous values, models those records by utilizing K-partite graph, and proposes clustering algorithm and matching algorithm to cope with duplicates and conflicting data.
2010.11.19 Venue: FL1, Meeting Room, Information Building	
Yun Deng (Web Group)	**Evaluating Entity Resolution Results** [ppt] Abstract: Entity Resolution is an important technique in data integration. Similar to clutering and partition, ER tries to identity the same entity among messes of records. This report focus on an ER results measure,GMD.
Jing Zhao (Cloud Group)	**Research on Query Processing** Abstract: Query Processing is an difficult problem in both parallel database and cloud-based database. We briefly introduce basic query processing steps in centralized database and parallel database, and talk something about web-scale query processing, including MapReduce debates, MapReduce-based join algorithms, etc. Finally, we introduce main idea of our work and some future work.
2010.11.14 Venue: FL1, Meeting Room, Information Building	
Lizhen Fu (XML Group)	**Diversification for Keyword Search on Graph Data** Abstract: Keyword search is the de facto information retrieval mechanism for data on theWorld WideWeb. It also proves to be an effective mechanism for querying semi-structured and structured data, because of its user friendly query interface.Recently, query processing over graph-structured data has attracted increasing attention.In this report,we focus on the semantic Diversification of results from keyword search on graph.
Qingling Cao (Flash Group)	**Enterprise Application of SSD** [ppt] Abstract: SSD is becoming more and more popular in enterprise.But there is a

	question,if the platform ready for SSD？This report solved the question.And it also introduced about SSD RAID.
2010.11.06 Venue: FL1, Meeting Room, Information Building	
Yingjie Shi (Cloud Group)	**CIKM2010 Story** Abstract: In this talk, I presented some papers and one panel related to Cloud Data Management in CIKM2010. Then I gave some summary of CIKM2010.
Bingbing Liu (Cloud Group)	**RHP:a new partitioner to improve the efficiency of range query in cassandra** Abstract: The conflicting problems of ensuring data-access load balancing and efficiently processing range queries leads to that cassandra can't support range query very well.So how to trade off them is the key point.
2010.10.30 Venue: FL1, Meeting Room, Information Building	
Dongqi Liu (Mobile Group)	**Spatial-temporal sequence views query demo** [ppt] Abstract: We have taken some informations of views on flicker to analyse how to traverse these views from the realistic perspective.If a user wants to traverse the views in a limited time,he may have several solutions,but which one is the most valuable one?Based on our ideas,we give three solutions to slove this problem,and will show you the solutions in our demo.
Long Liu (Cloud Group)	**Survey of Object-based Storage** [ppt] Abstract: Object-based Storage, a new approach to storage technology, is a subject of academic research and development in the storage industry. This survey describes the main points of object-based storage technology from five aspects. That is why we introduce the concept of object-based storage, what it is, how to take advantage of it, what the status of object-based storage in both industry and academic research is, and what we can do about it.
Yi Huang (Mobile Group)	**Android Development tutorial** [ppt] Abstract: Android, released by Google on Nov. 5th, 2007, is a Linux kernel-based operating system designed for smartphones. In the past three years, Android system has archived a great market share and this share is still increasing. Meanwhile, Android has been attracting more and more developers who have made contributions to more than 100,000 applications in the second largest online app store called Android Market. This tutorial introduces application development on Android platform and the mechanism of Android as well.
2010.10.23 Venue: FL1, Meeting Room, Information Building	

Fan Yulei (Mobile Group)	**Flash-based Multi-Version Data Storage** Abstract: Because of characteristics of Flash Memory and Data storage of PostgreSQL, More update operations and small random write operations run on flash memory. These operations will degrade the performance of DBMS and age of flash memory. Flash-based Multi-Version Data Storage(FMVDS) is proposed to reduce update and write operations and finally reduce erase times. In FMVDS, transaction table item with timestamp and data record with a point to older version data implement high concurrency control and quickly recovery.
Daxing Jiang (MSRA)	**Context-Aware Search** Abstract: Introduce the research on context-aware search in MSRA.
2010.09.25 Venue: FL1, Meeting Room, Information Building	
Youzhong MA (Web Group)	**Entity Resolution with Evolving Rules** [ppt] Abstract: Entity resolution (ER) identifies database records that refer to the same real world entity. In practice,ER is not a one-time process,but is constantly improved as the data, schema and application are better understood. We address the problem of keeping the ER result up-to-date when the ER logic ñevolvesò frequently. A naive approach that re-runs ER from scratch may not be tolerable for resolving large datasets. This paper investigates when and how we can instead exploit previous ñmaterializedò ER results to save redundant work with evolved logic. We introduce algorithm properties that facilitate evolution, and we propose efficient rule evolution techniques for two clustering ER models: match-based clustering and distance-based clustering. Using real data sets, we illustrate the cost of materializations and the potential gains over the naive approach.
Jinzeng Zhang (Mobile Group)	**VLDB paper report** Abstract: This report includes two parts.The fisrt is retrieving top-k prestige-based relevant spatial web objects,this method proposes the concept of prestige-based relevance, the top-k spatial web objects is ranked according to both prestige-based relevance and location proximity.The second part introduces how to mine significant sematic location from GPS data,this method models the relationships between locations and the relationships between locations and users with a two-layered graph.Based on this,this paper proposes a new ranking model which assign significance to locations.
Yingjie Shi (Web Group)	**Paper Summary of VLDB2010** Abstract: Papers of VLDB2010 about cloud are classified into four aspects:

	Cloud Data Management Systems, Benchmark, Query Processing and open questions. This report introduces the motivation, key technology and inspiration to our research work.
2010.09.18 Venue: FL1, Meeting Room, Information Building	
Zhongyun Wang (Graduate)	**New Experience in MSRA** Abstract: Introduce personal life , feelings in MSRA.
Da Zhou (Graduate)	**Introduction to Cloud and Flash Memory Management** Abstract: Share new findings and thoughts about cloud computing and flash memory management.
2010.06.19 Venue: FL1, Meeting Room, Information Building	
Zheng Huo (Mobile Group)	**Privacy-preserving of Trajectory Data: A Survey** [ppt] Abstract: This survey discussed trajectory data privacy preservation techniques in 4 motivating applications. For online trajectory data privacy preservation, service is centric, trade-off is between QoS and privacy preservation; For offline trajectory data privacy preservation, data is centric, trade-off is between data quality and privacy preservation.
Qingsong Guo (XML Group)	**XML Keyword Query Refinement** [ppt] Abstract: In this report, we discussed about the problem of query refinement in traditional IR and novel XML keyword search. The main part we mentioned is about the task and ways of XML keywords query refinement. In addition, we classified the existing work of XML keywords query refinement, and give out my own work on it.
2010.06.12 Venue: FL1, Meeting Room, Information Building	
Ruxia Ma (Web Group)	**Credibility on the Web: A Survey** Abstract: This survey discussed credibility on the web from three kinds of entities
Wei Chen (Web Group)	**Information Quality and Trustworthiness in Wikipedia** Abstract: In this talk we discussed the problem of information quality and trustworthiness of Wikipedia and introduced some research topics. In addition, we gave an brief overview of current research papers about this topic in WWW, WICOW etc.
2010.06.05 Venue: FL1, Meeting Room, Information Building	

Xiangmei Hu (Cloud Group)	**Index for cloud data management** Abstract: This report mainly introduces why we build index on cloud data management、some related work about index for cloud data management and our work progress on index research.
Haiping Wang (Cloud Computing Group)	**NoSQL Overview** [ppt] Abstract: This report simply introduced NoSQL,four reasons why nosql concept was introduced, the history, definition,Three fundamental theories of NoSQL and categories of NoSQL databases.
2010.05.29 Venue: FL1, Meeting Room, Information Building	
Lizhen Fu (XML Group)	**Keyword search on Graph** Abstract: In this report, I introduce methods that perform keyword search on graph data. Keyword search provides a simple but user-friendly interface to retrieve information from complicated data structures. In this discussion, I focus on three major challenges of keyword search on graphs. First, an answer to a keyword search on graphs,or, what qualifies as an answer to a keyword search. second, what constitutes a good answer, or how to rank the answers;Third, how to perform keyword search efficiently.
Lizhen Fu (XML Group)	**The Integration of TelecommuniCations Networks, Cable TV Networks and The Internet** [ppt] Abstract: This report introduces the conception The Integration of TelecommuniCations Networks, Cable TV Networks and The Internet firstly.then present its development Process and its advantages. At last,I describe the current situation of Integration of the three kides of networks at abroad.
2010.05.22 Venue: FL1, Meeting Room, Information Building	
Yubo Kou (Web Group)	**Elementary Structure-based Graph Matching** Abstract: Past graph matching techniques is vertex-based. Which means they first find candidate set for each node in the query, then perform searching algorithm to find a match. This approach cost too much since there might be too many candidates for each node, and these candidates will form a large search space. To reduce the search space, it is profitable to elevate the granularity of matching algorithm
Wei Wang (XML Group)	**Data deduplication** Abstract: This report introduces some methods of data deduplication, such as Hash-based algorithms, Delta algorithms.

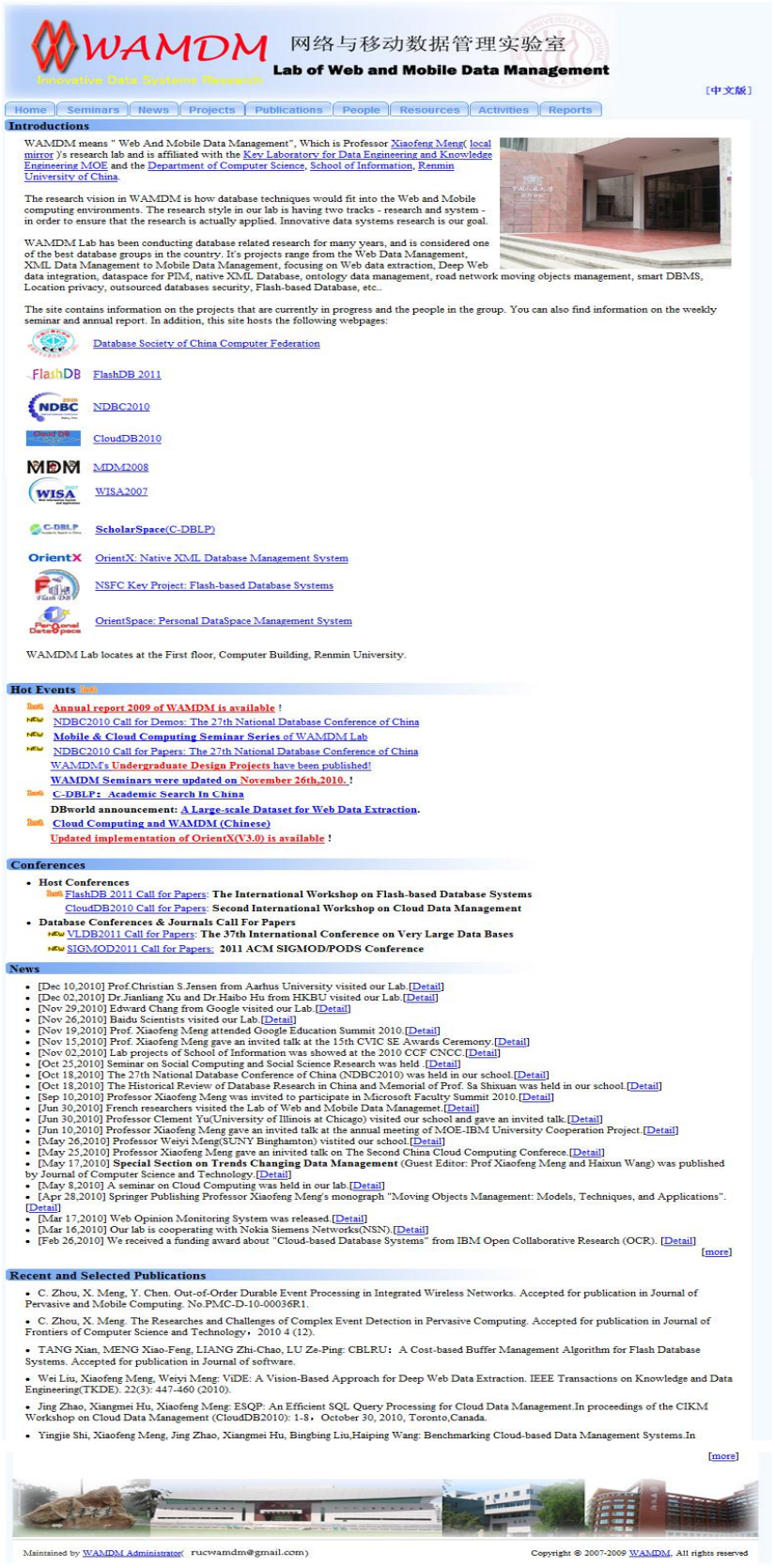
2010.05.08 Venue: FL1, Meeting Room, Information Building	
Yingjie Shi (Web Group)	**Benchmark results and analysis** Abstract: This report introduces the test results of benmarks on cloud-based DBMSs, and does analysis on the restuls.
Haiping Wang (Cloud Computing group)	**Architecture and Design of Distributed Database Systems** [ppt] Abstract: This report introduces serval kinds of architectures about Distributed Database Systems based on relational data model, it also introduces two horizonal and a verical fragmentatin method and the allocation model for DDBMS.
2010.04.24 Venue: FL1, Meeting Room, Information Building	
Xuan Zhou (CSIRO, Australia)	**Integrating User Interfaces of DB and IR Systems** Abstract: In contrast to classical databases and IR systems, real-world information systems have to deal increasingly with very vague and diverse data structures. While current object-relational database systems require clear and unified data schemas, IR systems usually ignore the structured information completely. Malleable schemas, as recently introduced, provide a novel way to deal with vagueness,ambiguity and diversity by incorporating imprecise and overlapping definitions of data structures. In this talk, I will introduce a novel query relaxation scheme that enables users to find best matching information by exploiting malleable schemas. Our scheme utilizes duplicates to discover the correlations within a malleable schema, and then uses these correlations to appropriately relax users' queries.Then, it ranks results of the relaxed queries according to their respective probability of satisfying the original queryõs intent. Our experiments with real-world data confirmed its performance and practicality.
2010.04.17 Venue: FL1, Meeting Room, Information Building	
Zhichao Liang (Flash Group)	**Hush-Tell You Something Novel About Flash Memory !** Abstract: This report introduces some work of Non-volatile Systems Laboratory in UCSD in which a lot of tests on flash memory were done. According to the test results, some applications were deviced, including a variation-aware FTL which is called Mango, a flash-aware data encoding and a system architecture for data-centric applications whose name is Gordon.
Yulei Fan (Mobile Group)	**Existed DBMS on SSD** Abstract: By analysis of IOps of HDD and SSD,we can compare IOps of SSD with IOps of HDD. By analysis of tpcc of MySQL and PG on SSD and HDD, we can compare performance of existing DBMS on SSD

	with that on HDD. Then we propose some ideas
2010.04.03 Venue: FL1, Meeting Room, Information Building	
Zhongyuan Wang (Web Group)	**Web Pages Extraction Technologies in the Opinion Monitoring System** Abstract: This report introduces two web pages extraction technologies in our opinion monitoring system, and some popular tools for system development.
Yi Huang (Mobile Group)	**An Introduction to Flex** [ppt] Abstract: Nowadays Flex is very popular in developing Rich Internet Applications. This report introduces what is Flex and its history and also discusses its mechanism, advantages, applications and the differences between other RIA techniques.
Jing Zhao (Web Group)	**System Environment and MapReduce Framework** Abstract: This report includes the introduction of the construction of our cloud data management platform and a brief talk about MapReduce framework.
Zhichao Liang (Flash Group)	**An Introduction to the Source Insight** [ppt] Abstract: This report introduces a project-oriented program editor and code browser,Source Insight,which parsers your source code and maintains its own database of symbolic information dynamically while you work,and presents useful contextual information to you automatically.
2010.03.27 Venue: FL1, Meeting Room, Information Building	
Chunjie Zhou (Web Group)	**IO3:Interval-based Out-of-order Event Processing in Pervasive Computing** Abstract: In pervasive computing environments, complex event processing has become increasingly important in modern applications. A key aspect of complex event processing is to extract patterns from event streams to make informed decisions in real-time. However, network latencies and machine failures may cause events to arrive out-of-order. In addition, existing literatures assume that events do not have any duration, but events in many real world application have durations, and the relationships among these events are often complex. In this work, we first analyze the preliminaries of time semantics and propose a model of it. A hybrid solution including time-interval to solve out-of-order events is also introduced, which can switch from one level of output correctness to another based on real time. The experimental study demonstrates the effectiveness of our approach.
Bingbing Liu (Cloud Group)	**ICDE2010 Keynote - what's new in the cloud** [ppt] Abstract:

	This report talks about why we should do cloud computing,how to do and what to do.
Yukun Li (Web Group)	**Survey of ICDE2010 and SIGMOD2010** Abstract: Based on the accepted papers, this presentation made a survey on recent international database conferences ICDE2010 and SIGMOD2010, and analyzed the research focuses of database area.
2010.03.20 Venue: FL1, Meeting Room, Information Building	
Da Zhou (Flash Group)	**RWConvertor: Random Write Optimization for SSD** Abstract: With the development of electronic technologies, Solid State Drive (SSD) emerge as new data storage media with low power consumption, high shock resistance and lightweight form. Besides these, the most attractive characteristic is the high random read speed because of no mechanical latency. Therefore SSD have been widely used in laptops, desktops, and data servers in place of hard disk during the past few years. However, poor random write performance becomes the bottle neck in wider applications. Random write is almost two orders of magnitude slower than both random read and sequential access, so write-intensive applications have very low performance on SSD. In this paper, the first time we propose to insert unmodified data into random write sequence in order to convert random writes into sequential writes, and then data sequence can be flushed at the speed of sequential write. Further, we improve the write performance by Optimum Converted Write Sequence (OCWS). Strict mathematical proof decides the location and number of inserted data items during the course of getting OCWS. We also optimized our method with throughput, which is decided by gain and granularity, of OCWS when applied in data stream.
2010.03.13 Venue: FL1, Meeting Room, Information Building	
Jinzeng zhang (XML Group)	**Approaches to internet of things** Abstract: As the next generation of information technology,the internet of things has drawn public attenention.It enables the internet to reach out into the real world of physical objects.This report first gives the concept of the internet of things,then introduces the system architecture and key techniques and gives three applications.Fianlly,I put forward to the furture direction.
Xing Hao (Mobile Group)	**Related Work about Internet of Things** [ppt] Abstract: This report gives an overview of the related and future work about Internet of Things and focus on the The RFID Ecosystem Experience handled by University of Washington.
2010.03.06 Venue: FL1, Meeting Room, Information Building	

Yingjie Shi (Web Group)	**Open Source Cloud-based DBMS Experiments** Abstract: This report introduces existing expriment benchmarks of cloud-based DBMS experiments. We describe the testbed of our experiment, and show the tasks and results.
Zhongyuan Wang (Web Group)	**System Architecture Design and Implementation of Cloud-based Database System** Abstract: The Cloud-based Database project at WAMDM aims at researching new storage and database system which can support the next generation of data storage and management and applied to mobile communications. This report introduced the architecture design and implementation of our cloud-based database system.
2010.01.09 Venue: FL1, Meeting Room, Information Building	
Dr. Yueguo Chen (Invited Talk)	**Time series and Interactive media** Abstract: Time series and interractive media have large applications in computer games or so. One of the most important problem for pattern detection in streaming time series could be how to define a effective distance metric.We propose a novel warping distance and efficient approach for continuous pattern detection. For the interavtive media database, it focus on the index,storage structure for smart media objects, similarity metrics and query procesing on multimedia data.
Xiaoying Qi (Flash Group)	**FTL Algorithms and Native Flash Experiments** Abstract: This report introduces five flash translation layer algorithms, such as BAST, FAST, LAST, and DFTL etc. We mainly describe the main ideas of those algorithms and their realization. Then we introduce the native flash experiments.

实验室网站



http://idke.ruc.edu.cn/wamdm

实验室成员

Faculty Members

Xiaofeng Meng
孟小峰
博士，教授，博导
WAMDM 实验室负责人

Nan Yang
杨楠
博士后，副教授

Qing Liu
刘青
博士，副教授

Yunpeng Cai
柴云鹏
博士，讲师

Gang Yang
杨刚
博士，讲师

Zhiyong Shan
单智勇
博士，讲师

Ph.D. Candidates

Chunjie Zhou
周春姐

Xian Tang
汤显

Yulei Fan
范玉雷

Lizhen Fu
富丽贞

Zheng Huo
霍峥

Jinzeng Zhang
张金增

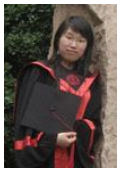
Yingjie Shi
史英杰

Ruxia Ma
马如霞

Xiaojian Zhang
张啸剑

Youzhong Ma
马友忠

M.Sc. Students



Jing Zhao
赵婧

Xiangmei Hu
胡享梅

Wei Wang
王伟

Qingsong Guo
郭青松

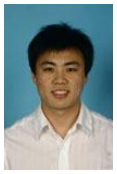
Zeping Lu
卢泽萍

Zhichao Liang
梁智超

Xiaoying Qi
綦晓颖

Yi Huang
黄毅

Haiping Wang
王海平



Bingbing Liu
刘兵兵

Wei Chen
陈威

Jie Wen
文洁

Dongqi Liu
刘东琦

Long Liu
刘龙

Qingling Cao
曹庆铃

Wei Tong
童薇

Yun Deng
邓云

Sen Yang
杨森

实验室毕业生

2010 年毕业生去向

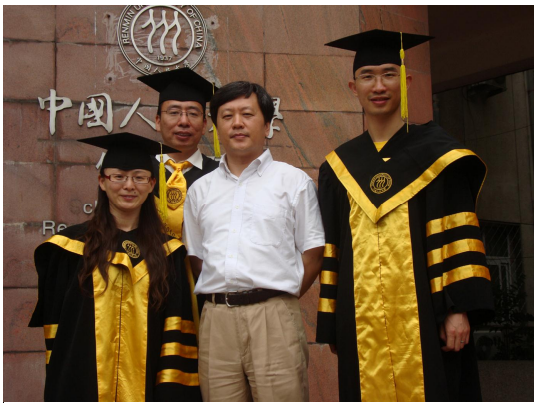
姓名	学历	时间	毕业去向
李玉坤	博士	2010 年 7 月	天津理工大学
潘晓	博士	2010 年 7 月	石家庄铁道大学
周大	博士	2010 年 7 月	中国移动研究院
徐俊劲	硕士	2010 年 7 月	百度
王仲远	硕士	2010 年 7 月	微软亚洲研究院
艾静	硕士	2010 年 7 月	国家发改委国家投资项目评审中心
郝兴	硕士	2010 年 7 月	百度
张相於	硕士	2010 年 7 月	搜狗

2009 年毕业生去向

姓名	学历	时间	毕业去向
周军锋	博士	2009 年 7 月	燕山大学
姜芳芃	博士	2009 年 7 月	徐州师范大学
贾琳琳	硕士	2009 年 7 月	中国农业银行
黄静	硕士	2009 年 7 月	中国工商银行软件开发中心
朱金清	硕士	2009 年 7 月	百度
王伟	硕士	2009 年 7 月	百度
向锂	硕士	2009 年 7 月	中化集团石油中心

实验室毕业生集体照片

2010 年毕业生



2009 年毕业生

2005 年毕业生