实验室研讨会

2009.12.26 Venue: FL1, Meeting Room, Information Building	
Rui Zhang	**Continuous Intersection Join on Moving Object** Abstract: The continuous intersection join query is computationally expensive yet important for various applications on moving objects. No previous study has specifically addressed this query type. We can adopt a naive algorithm or extend an existing technique (TP-Join) to process the query. However, they compute the answer for either too long or too short a time interval, which results in either a very large computation cost per object update or too frequent answer updates, respectively. This motivates us to optimize the query processing in the time dimension. In this study, we achieve this optimization by introducing the new concept of time-constrained (TC) processing.
Jinchuan Chen	**Data Integration with Uncertainty** Abstract: A survey of data integration with uncertainty. This report introduced existing methods of data management with uncertainty.
2009.12.19 Venue: FL1, Meeting Room, Information Building	
Haiping Wang (Cloud Computing Group)	**cassandra and sigmod contest** Abstract: Cassandra is a highly scalable second-generation distributed database, bringing together Dynamo's fully distributed design and Bigtable's ColumnFamily-based data model.The task of sigmod programing contest 2010 is to implenment a simple distributed query executor built on top of the last year's main-memory index.
Ying Lu (Mobile Group)	**Hammer & Nail** Abstract: "Research is actually a process of hammers(methods) hammer nails(problem)". This report first presents three hammers, i.e.three kinds of hash functions, which are signature, OPMPHF(Order Preserving Minimal Perfect Hash Function) and LSH(Location Sensitive Hashing).Then it introduces a nail using the hammers above.It is called Reveser k Spatial and Textual Nearest Neighbor(RkSTNN).
2009.12.12 Venue: FL1, Meeting Room, Information Building	
Yingjie Shi (Web Group)	**Survey on Data Management in the Cloud** Abstract:

	With the development of computer and communication technology, a large scale of data are produced. Cloud-based database is one solution to efficiently store and analyze these data. In this talk, we present some cloud-based database and summarize them from different aspects.
Bingbing Liu (Cloud Computing Group)	**Hive – A Warehousing Solution Over a MapReduce Framework** Abstract: Introduce a system which support managing and querying structured data and builded on the top of hadoop and the query language.
2009.12.05 Venue: FL1, Meeting Room, Information Building	
Ruxia Ma (Web Group)	**Trust Metric on Social Network** Abstract: This report introduces five trust metric mechanisms on social network, such as: Advogato, Appleseed and TidalTrust, etc. We mainly describe the main ideas of those algorithms and their realization.
Wei Chen (Web Group)	**Data Fusion-Resolve Data Conflicts in Integration** Abstract: In this talk we gave a brief introdution to data fusion, including data conflict types, conflict resolution strategies, the role played by data fusion in integration programs and current approaches to data fusion. Then we addressed some challenges and open problems in data fusion research. Finally we presented a brief summary to this talk.
2009.11.28 Venue: FL1, Meeting Room, Information Building	
Xian Tang (mobile Group)	**ACR: an Adaptive Cost-Aware Buffer Replacement Algorithm for Flash Storage Devices** Abstract: In this talk, we propose an adaptive cost-aware buffer replacement algorithm--ACR, which adapt to various access patterns on flash disks.
Yulei Fan (Mobile Group)	**Multi-version Concurrency Control of Database Based on Flash Memory** Abstract: Data may have multiple versions as because of the feature of not-in-place update and in-page logging store mechnism in flash memory. Multi-version concurrency control has to be implented based on the Serialization theory, and it includes MV2PL(multi-version 2PL), MVTO(multi-version TO), MVSGT(multi-version SGT), TW(time warp) and ROMV(read-only multi-version). We evaluated the performance of these algorithms by implementing experiments on existing DBMS such as MS SQLServer, MySQL and Postgres. Finally,

	we proposed some future work in Multiple-version Concurrency Control.
2009.11.21 Venue: FL1, Meeting Room, Information Building	
Junjing Xu (XML Group)	**Efficient String Similarity Search Using Synonyms** Abstract: This report introduces the gram_based string matching functions and the new similarity funcion.
Lizhen Fu (XML Group)	**Reachability Queries on Large Directed Acyclic Graphs** **Abstract:** In particular, graph reachability has attracted a lot of research attention as reachability queries are not only common on graph databases, but they also serve as fundamental operations for many other graph queries. In this reprot, I introduce my new graph label to speed up the processing of reachablity queries on DAG， which index is small and which can be constructed easily
Jinzeng Zhang(XML Group)	**Information Retrieval Model and Relevance Feedback** **Abstract:** This report first introduces four classic information retrieval models. Based on those models, we present two methods of improving retrieval results
2009.11.14 Venue: FL1, Meeting Room, Information Building	
Yukun Li(Web Group)	**Review our studies on dataspace** **Abstract:** Reviewed our works on dataspace research, and introduced a work we are doing.
Xiangyu Zhang(Web Group)	**Dataspace Research Report** Abstract: Introduced research and system implementation progress on Dataspace research.
Yubo Kou(Web Group)	**Leveraging Feature Context to Facilitate Sub-graph Query in Graph Database** Abstract: Previous techniques focus on feature selection strategy to filter false graphs as more as possible. This approach has met a bottleneck, that as the feature is becoming more and more complicated, precision is still low. Thus we propose to investigate into how feature context could help improve pruning power in sub-graph query.

2009.11.08 Venue: FL1, Meeting Room, Information Building	
Yukun Li(Web Group)	**About CIKM2009 Story** Abstract: Give a short summary on CIKM 2009 based on my impression on this conference, especially introduced the three keynotes.
Da Zhou(Mobile Group)	**Review of CIKM 2009** Abstract: CIKM is a high level international conference. There are three tracks
Zhongyuan Wang(Web Group)	**Summary of CIKM2009** Abstract: In this talk, I presented three papers and one tutorial related to Web data management and click log mining in CIKM2009. Then give some summary of CIKM2009.
Xiangyu Zhang(Web Group)	**IR is Interesting-CIKM 2009 Report** Abstract: In this presentation, I gave a brief summary and introduction to the CIKM 2009 conference and some of my own experience on this conference.
2009.10.31 Venue: FL1, Meeting Room, Information Building	
Xiangyu Zhang(Web Group)	**An Efficient Multi-Dimensional Index for Cloud Data Management** Abstract: In this presentation, I introduced our work of multi-dimensional index structure for Cloud Computing platforms.
Yukun Li (Web Group)	**Supporting Context-based Query in Personal DataSpace** Abstract: Many users need to refer to content in existing files (pictures,tables, emails, web pages and etc.) when they write documents(programs, presentations, proposals and etc.), and often need to revisit these referenced files for review, revision or reconfirmation. In this paper, we propose an efficient solution for this problem. We firstly define a new personal data relationship
Da Zhou(Mobile Group)	**Pre-Report for CIKM 2009** Abstract: Solid State Drive (SSD), emerging as new data storage media with high random read speed, has been widely used in laptops, desktops, and data servers to replace hard disk during the past few years. However, poor random write performance becomes the bottle neck in practice. In this

	paper, we propose to insert unmodified data into random write sequence in order to convert random writes into sequential writes, and thus data sequence can be flushed at the speed of sequential write.
2009.10.24 Venue: FL1, Meeting Room, Information Building	
Xiangye Xiao (Web&Mobile Group)	**Overview of Talks in NDBC 2009** Abstract: Dr. Xiangye Xiao gave a brief review of invited talks in NDBC 2009 which includes Dr. Xin Dong from AT&T, Prof. Weiyi Meng from Binghamton Univ., Haixun Wang from MSRA and Lei Chen from HKUST.
Yukun Li(Web Group)	**Report on SKG2009** Abstract: Give an introduction on SKG2009, and focusing on introducing the two keynotes of this conference.
Zheng Huo(Mobile Group)	**A new topic: queries with geo-information** Abstract: Discovering users' specific and implicit geographic intention in web search can greatly help satisfy users' information needs. Research on queries with geo-information has becoming hot these years. There are several methods. First, the training data based methods, these methods need big data of query logs; another is spatial and texual information retrieval methods, but these methods can only deal with local geo-informaiton. The challege is how to discover users' implicit geo-information in queries.
Xiangmei Hu(Web Group)	**Trajectory pattern mining** Abstract: The pervasiveness of mobile devices and location based services is leading to an incresing volume of mobility data.This side effect provides the opportunity to analyse the behaviors of movements.With this background,trajectory pattern mining has been a popular topic.This report mainly introduces some representative work about this topic and points out some defects.
2009.10.11 Venue: FL1, Meeting Room, Information Building	
Jing Ai(Web Group)	**C-Rank -- A Credibility Evaluation Method for Deep Web Records** Abstract: How to identify and evaluate information credibility ranking has become an increasing important problem. To address the issue, an effective credibility evaluation method called C-Rank to compute trust

	values of records in Deep Web databases is proposed, which constructs an S-R Credibility Graph for each record.
Xing Hao(Mobile Group)	**Privacy Preserving towards Continuous Query in Location-based Services** Abstract: With advances in wireless communication and mobile positioning technologies, location-based mobile services have been gaining increasingly popularity in recent years. Privacy preservation, including location privacy and query privacy, has recently received considerable attention for location-based mobile services. A lot of location cloaking approaches have been proposed for protecting the location privacy of mobile users. However, they mostly focus on anonymizing snapshot queries based on proximity of locations at query issued time. Therefore, most of them are ill-suited for continuous queries. In view of the privacy disclosure (including location and query privacy) and poor quality of service under continuous query anonymization, a δp-privacy model and a δq-distortion model is proposed to balance the tradeoff between privacy preserving and quality of service. Meanwhile a temporal distortion model is proposed to measure location information loss during a time interval, and it is mapped to a temporal similar distance between two queries. Finally, a greedy cloaking algorithm (GCA) is proposed, which is applicable for both anonymizing snapshot queries and continuous queries. Average cloaking success rate, cloaking time, processing time and anonymization cost for successful requests is evaluated with increasing privacy level (k). Experimental results validate the efficiency and effectiveness of the proposed algorithm.
Wei Wang(XML Group)	**Algebra-based Transform query optimization strategy** **Abstract:** XQuery/Update defines a special Transform query, which is similar to be hypothetical query in relation databases, and can be expressed as“Q when {U}”. In other words, the results of query Q are the same as the results after executing hypothetical update {U} on the original database, without actually updating database. The Transform queries need to copy the nodes in XML database and then update copied nodes, so it doesn’t affects the database. But Transform queries will usually copy and update a lot of nodes which are useless for query Q and result in high cost. It is critical for query optimization to decrease the number of copied nodes and the update operation. In this paper, we propose a set of rules for Transform query optimization techniques based on OrientXA. Which are implemented in OrientX3.0.
Da Zhou(Mobile	**HF-Tree--An Update-Efficient Index for Flash Memory**

Group)	Abstract: Due to the expensive write cost of flash memory, traditional disk-based indexes have a poor update performance when directly applied to flash drives. In this talk, Da Zhou proposed a novel index called HF tree to improve the update performance of Flash memory, which integrates BF -tree with Tri-hash.
Zhichao Liang(Mobile Group)	**Sub-Join--A Query Optimization Algorithm for Flash-based Database** Abstract: Compared with Hard Drive Disk (HDD), SSD has a lot of advantages, such as high random read performance, low power consumption and lightweight form. Therefore it is envisioned to be next generation data storage instead of HDD. However, the enhancement of query performance for flash-based database is not the same as the IO ratio of SSD to HDD. The reason is existing databases which are designed for HDD can not take full advantage of high IO performance of SSD. In this paper, a new join algorithm, Sub-Join, is proposed. Sub-Join first projects the column of join and primary key as Sub-Table, and then executes join operations on Sub-Tables. Finally results are gotten from original table according to the result of join on Sub-Tables. The compared experiments with Oracle Berkeley DB show Sub-Join outperforms original indexed nested-loop join at the ratio of about 40%~100%. The result strongly shows the high efficiency of this method.
2009.09.28 Venue: FL1, Meeting Room, Information Building	
Xin(Luna) Dong(AT&T Research)	**Data Integration with Uncertainty** Abstract: Dr. Xin (Luna) Dong from Data Management Department at AT&T Research visited Web And Mobile Data Management (WAMDM) lab and gave an invited talk about Data Integration with Uncertainty. Her talk mainly focused on some important and valuable topics in uncertain data integration.
Xiangye Xiao(Web&Mobie Group)	**Efficient Co-Location Pattern Discovery** Abstract: Dr. Xiangye Xiao gave a brief talk about her research topics when she was a PHD candidate in the Hong Kong University of Science and Technology. Her talk included efficient co-location pattern discovery and Web browsing on mobile devices. Besides, Dr. Xiangye Xiao proposed some ideas about future research.

Jiaheng Lu(XML Group)	**Keyword Search Techniques in Mobile Web** Abstract: Dr. Jiaheng Lu received an a funding award about "keyword search in mobile web" from National Science Foundation China (NSFC). He gave a detailed demonstration about the project and proposed some possible topics.
2009.07.25 Venue: FL1, Meeting Room, Information Building	
Qingsong Guo(XML Group)	**OrientX4.0 - Supproting Keyword Search** Abstract: With the developing of xml technology, more and more pepole using xml data. In traditional, we use the standard query lanaguage XQuery to find the data we need, but we need to learn the "XQuery" and we must know the structure and content of the xml document. It is great challenge of naive users. For this popose, in the new edition-OrientX4.0, we supporting the xml keyword-search , which can solve the problem we meet by using XQuery and make pepole using xml more easier.
Wei Wang(XML Group)	**OrientX4.0 System Development Report** Abstract: the implement of XML keyword search
2009.07.18 Venue: FL1, Meeting Room, Information Building	
Xing Hao(Mobile Group)	**Probabilistic kNN Query in Road Network** Abstract: Queries for moving objects in road network, especially kNN(k Nearest Neighbor) queries are very important and have received considerable attention. This speech discusses how to model the uncertainty data and process kNN queries in road network.
Yi Huang(Mobile Group)	**Report on Privacy Protection Demo Appplication Development** Abstract: In order to apply the current privacy protection algorithms and integrate them in the 863 Pervasive Computing project, we decided to develop a demonstration application.This report introduced the technical and functional characteristics of the application as well as the development plan.
Chunjie Zhou(Mobile Group)	**Query Processing over Interval-based Out-of-order Event Streams** Abstract: Complex event processing has become increasingly important in modern applications, ranging from supply chain management for RFID

	tracking to real-time intrusion detection. 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 at the event processing engine. In addition, existing temporal pattern mining assumes that events do not have any duration. However, events in many real world applications have durations, and the relationships among these events are often complex. In this work, we propose solution to process both sequence and parallel pattern queries on out-of-order event streams. First, we analyze the preliminaries and the problems caused by out-of-order data arrival. We then propose a method to detect out-of-order event patterns. A new solution including time-interval to solve out-of-order problems is also introduced. Lastly, we conduct an experimental study demonstrating the effectiveness of our approach.
2009.07.11 Venue: FL1, Meeting Room, Information Building	
Zhichao Liang(Mobile Group)	**System Development Report of Flash Group** Abstract: Our target is to develop a special flash-based DBMS,and we decide to do some changes on an existing open source DBMS to work it out. However, as a matter of fact,there are lots of open source systems. Which one is the best choice? After a detailed analysis, we believe MySQL,which contains the Berkeley DB as one of its storage engines,is the answer to our problem.
2009.07.04 Venue: FL1, Meeting Room, Information Building	
Yukun Li (Web Group)	**SIGMOD2009 Overview** Abstract: Analyze the current hot research issues based on the accessed papers of SIGMOD2009, and introduce two papers of this conference.
Da Zhou (Mobile Group)	**Flash Research Report** Abstract: Flash-based database systems research becomes more and more hot. In sigmod2009 and VLDB2009, we are glad to see that there are some papers about the indexing, query processing and transaction processing. This report gives a coarse overview to the motivations and ideas of these papers.
Lizhen Fu (XML Group)	**XML Labeling and Query Optimization in Sigmod09** Abstract: Optimization of complex XQueries combining many XPath steps and joins is currently hindered by the absence of good cardinality

	estimation and cost models for XQuery.Labeling schemes lie at the core of query processing for many XML database management systems. Designing labeling schemes for dynamic XML documents is an important problem that has received a lot of research attention. This presention introduce a new labeling scheme DDE and a new Runtime Optimization approach ROX in sigmod09.
2009.06.27 Venue: FL1, Meeting Room, Information Building	
Zeping Lu (Mobile Group)	**Logging in Flash-based Database Systems** Abstract: Synchronous transactional logging is the central mechanism for ensuring data persistency and recoverability in database systems. In this report,we discussed the solutions about exploiting different kinds of flash drives for synchronous logging and the recovery processing technologies related with them.
Xiangmei Hu (Web Group)	**Location-based Database Selection** Abstract: Location_based database selection is a new topic,This report mainly gives an introducton about this topic,including why we choose this topic,what the problem is,some related work and how to solve the problem.
Jing Zhao (Web Group)	**Snippet of Structured Data** Abstract: It is expected that more and more people will search the web when they are on the move. But there are many limitations when we browsing the web page in mobile devices, especially small screen. A record in database usually contain lots of information, which is not useful for user and is so much for small screen. So we try to extract the most useful attributes to return to user.
2009.06.20 Venue: FL1, Meeting Room, Information Building	
Qingsong Guo Wei Wang(XML Group)	**XML Keyword-Search engine** Abstract: XML has already became the de-facto of data exchange. So, how to query XML data is becoming very important. We can use the query language XQuery and XPath, which is the standard query language of XML recommended of W3C, to get what we need. But the user must be familiar with the query languages, and know the content and structure of XML data at first, so that the users can write the accurate query. It is not easy for most users, and it forcing the study of XML keyword-search, With it, we needn't learn the XML query language,

	and also, we needn't known the content and structure of XML. It make the query easier. The main features of next edition of OrientX(edition 4.0) is to supprot the keyword-search, in the presentation, qingsong guo analized the existing XML keyword-search engine and made a comparison and get their features in common . And based it, we defined the main features of OrientX 4.0 . Wei wang analized the key technologies of xml keyword-search, such as the priciple and algorithms of computing SLCA, the ranking of query results.
2009.06.13 Venue: FL1, Meeting Room, Information Building	
Lizhen Fu (XML)	**Query Processing over Graph-structured XML Data** Abstract: When XML documents are modeled as graphs, many research issues arise. In particular, there are many new challenges in query processing on graph-structured XML documents because traditional query processing techniques for tree-structured XML documents cannot be directly applied.
Yulei Fan (Mobile Group)	**MVCC on Flash Memory** Abstract: First, Flash has the characteristic of Out-of-Place Updating, which lead to multiple version of data on Flash. Second, I introduce the basic priciple and some protocols of MVCC, such as MVSR, MVCR, MVTO, MV2PL and so on. Finally, I present some information of transaction in BDB and PG.
2009.06.06 Venue: FL1, Meeting Room, Information Building	
Xiao Pan (Mobile Group)	**Location,Location, Location** Abstract: This talk focuses on the dicussion of Keynote of Christian S. Jensen on MDM2009.
Yukun Li (Web Group)	**C-Query: Context-based Query in Personal DataSpace** Abstract: Many users need to refer to content in existing files (pictures,tables, emails, web pages and etc.) when they write documents(programs, presentations, proposals and etc.), and often need to revisit the referenced files for review, revision or reconfirmation. In this paper, we propose an efficient method for users to revisit these refferenced files by identifying a context-based refference relationship.
2009.05.23 Venue: FL1, Meeting Room, Information Building	

WangWei GuoQingsong (XML Group)	**OrientX system development report** Abstract: The main features of OrientX3.5 version and its implementation.
2009.05.16 Venue: FL1, Meeting Room, Information Building	
Da zhou (Mobile Group)	**Random Write Optimization for SSD** Abstract: Random write of SSD has low IO performance when compared with sequential/random read and write. This paper propose a novel method to avoid the low performance of random write.
Xian Tang (Mobile Group)	**buffer management policy** Abstract: In this talk, I introduced several interesting buffer management algorithms, including some algorithms which work well on disk-based DBMS, others are buffer management algorithms on flash-based DBMS.
2009.04.25 Venue: FL1, Meeting Room, Information Building	
Zhongyuan Wang (Web Group)	**An Indexing Framework for Efficient Retrieval on the Cloud** Abstract: The emergence of the Cloud system has simplified the deployment of large-scale distributed systems for software vendors. The Cloud system provides a simple and unified interface between vendor and user, allowing vendors to focus more on the software itself rather than the underlying framework. Existing Cloud systems seek to improve performance by increasing parallelism. This paper explores an alternative solution, proposing an indexing framework for the Cloud system based on the structured overlay. Its indexing framework reduces the amount of data transferred inside the Cloud and facilitates the deployment of database back-end applications.
Xiangyu Zhang (Web Group)	**Data Management in the Cloud - Limitations and Opportunities** Abstract: Analysed data management applications that are suitable to move to the cloud platform and discussed remaining challenges of such movement.
2009.04.18 Venue: FL1, Meeting Room, Information Building	
Junfeng Zhou (XML Group)	**MCN: A New Semantics Towards Effective XML Keyword Search** Abstract: In this talk, We propose a new XML Keyword Search Semantics aiming at capturing meaningful results while avoiding returning meaningless

	results. This contribution is based on the observation that when talking about relationship between data elements, users query intension is always based on the relationship of real word entities.
Fangjiao Jiang (Web Group)	**Selectivity Estimation for Exclusive Query Translatio in Deep Web Data Integration** Abstract: In Deep Web data integration, some Web database interfaces express exclusive predicate,which permits only one predicate to be selected at a time. Accurately and efficiently estimating the selectivity of each Qe is of critical importance to optimal query translation. In this paper, we mainly focus on the selectivity estimation on infinite-value attribute which is more difficult than that on key attribute and categorical attribute. We start with two observations
2009.04.11 Venue: FL1, Meeting Room, Information Building	
Yukun Li (Web Group)	**Summary of ICDE2009 keynotes** Abstract: This slides give a summary on three keynotes of ICDE2009.
Da Zhou (mobile Group)	**ICDE 2009 Introduction** Abstract: ICDE is a very important international meeting about data management. In this conference, there are a lot of works related to flash-based database. transaction becomes an important topic in this field.
Zhichao Liang (Flash Group)	**Demo in ICDE 2009 Conference** Abstract: WEST(Web Entity Search Technologies),instead of returning webpages that are related to any people who happened to have the queried name,is to output a set of clusters of webpages,one cluster per each distinct person.Fa is a new system for automated diagnosis of system failures that is designed to address the SLO violations.UQLIPS is a Web-based integrated platform which performs online detection of near-duplicate occurrences over continuous video streams,as well as retrieval of near-duplicate clips from segmented video collections.
2009.04.04 Venue: FL1, Meeting Room, Information Building	
Xiao Pan (Mobile Group)	**Distortion-based Anonymity towards Continuous Query in Mobile Services** Abstract: Privacy preservation has recently received considerable attention for

	location-based mobile services. A lot of location cloaking approaches have been proposed for protecting the location privacy of mobile users. In this paper, we present continuous query privacy disclosed and worst QoS resulting from anonymizing continuous query.
Chunjie Zhou (Mobile Group)	**Complex Event Detection in Pervasive Computing** Abstract: In pervasive computing environments, wide deployment of sensor devices has generated an unprecedented volume of atomic events. However, most applications such as healthcare, surveillance and facility management, as well as environmental monitoring require such events to be filtered and correlated for complex event detection. Therefore how to extract interesting, useful and complex events from low-level atomic events is becoming more and more important in daily life. Due to the increasing importance of complex event detection, this paper proposes a framework of Complex Event Detection and Operation (CEDO) in pervasive computing. It gives an event model and extends current detection by incorporating temporal and spatial settings of events and different levels of granularity for event representation. We first show research issues, related works, and main research problems in this area. Then our current research works and the preliminary results are introduced. Finally, the research plan of my PhD project is presented for discussion.
2009.03.28 Venue: FL1, Meeting Room, Information Building	
Fangjiao Jiang (Web Group)	**Deep Web Integration:Querying Structured Data on the Deep Web** Abstract: In this report, I will introduce the background of Deep Web, the key technologies of Deep Web data integration and the active research groups. Then I will compare the metaquerier with metasearch engine. Finally I will give the research problems in the future.
Xiangmei Hu (Web Group)	**Database selection** Abstract: Database selection is a important topic,this report gives an introduciton to database selection and then introduces our new problem.
2009.03.21 Venue: FL1, Meeting Room, Information Building	
Yukun Li (Web Group)	**CoreSpace: A personal dataspace framework based on user activity** Abstract: Present a new framework of personal dataspace by hightlighting relationship between users and average objects, which provides more

	effective approaches of querying personal dataspace.
Yubo Kou (Web Group)	**An efficient method to Identify personal task** Abstract: Present a new method to identify personal task based on user access activity.
2009.03.14 Venue: FL1, Meeting Room, Information Building	
Fei Huang (Cloud Computing)	**Research Report on Map/Reduce Framework Based on Hadoop** Abstract: Map/Reduce is the crucial algorithm of Hadoop. It is a easy but powerful algorithm that can solve the problems based on mass data. In this report,I will introduce the concept of Hadoop and Map/Reduce, then the detail of how the Map/Reduce framework do jobs.
Yi Hu (Web Group)	**Introduction to HBase** Abstract: As sub-project of Hadooop, HBase focus on providing storage for the Hadoop Distributed Computing Environment. HBase is a table coloum-oriented operating. Its three-layer file system provides the feasible scheme for the distributing data storage while its three-layer architecture solves the problems of region assignment and region location. To get intuitionistic understanding of HBase, comparison with MySQL has been made in the test.
Wei Chen (Web Group)	**The Progress of C-DBLP's Development and Future Plans** Abstract: The develop team of C-DBLP system has added some attractive functions and features to the site based on user's feedback and researching demand since the release of C-DBLP. Besides, we are working on some interesting problems such as Name Disambiguation and Mining of Relations among Authors. This report presented the progress of C-DBLP's development and showed intuitive approaches to the research problems in C-DBLP. Also, we made a detailed plan for future work in C-DBLP.
2009.03.07 Venue: FL1, Meeting Room, Information Building	
Linlin Jia (Web Group)	**Study on Fast Approxmate Membership checking** Abstract: Introduce ISH for approximate membership checking and analyze its disadvantage. We propose a new index and a corrresponding algorithm, the experiments indicate that the new method is more efficient than ISH.

Junjin Xu (XML Group)	**String Similarity** Abstract: This report introduces the methods about counting string similarity, including edit distance and gram_based similarity.

2009.02.28 Venue: FL1, Meeting Room, Information Building

Jing Zhao (Web Group)	**Faceted Search** Abstract: A introduction to faceted search, including the evolution of faceted search, the differences between faceted search and navigational search, direct search, and differences between cluster, tag and facet.
Wei Chen (Web Group)	**Automatic Construction of Facet Hierarchies** Abstract: Facet hierarchies are the main forms of data organization in facet search system. They are used to support facet-based navigation and refine the search results through different facets. The construction of facet hierarchies is one of the most important research topics in facet search. Since most facet hierarchies in current systems are built mannually, the automatic construction method is in great need. This presentation addressed W. Dakka and P. G. Ipeirotis's research progress in automatic construction of facet hierarchies.

2009.01.11 Venue: FL1, Meeting Room, Information Building

Junfeng Zhou (XML Group)	**Survey of XML Database Technology** Abstract: In this talk, I give the main topics about XML database and explain the existing solutions using simple examples.
Lizhen Fu (XML Group)	**Graph DataBases** Abstract: This presentation introduces some rearch hotspots on Graph DataBases，including the construction of the index, the processing of containment queryquery and reachability query answering.