

学术报告

2003年12月2-4日
中国人民大学多媒体教学楼
敬请光临

Seminars on Selected Topics in Database Research

Speaker: Dr. Wang Haixun, IBM T. J. Watson Research Center

Abstract

The database area is experiencing a vigorous and often tumultuous growth as a result of a push by new technology combined with a pull by new applications. The goal of this seminar series is to provide a broad exposure to the research being done in this area. The series will include such topics as database indexing, XML, data mining, database language, and the stream DBMS. I will present, for each topic, a brief survey, some specific research issues and their tentative solutions. The series is also aimed at inspiring students to engage in original research in this active area.

Seminars 1: Indexing Techniques for Advanced Database Applications

Time: 2:00-4:30pm, Dec 3, 2003 Classroom: **4505#**

Indexing is one of the most fundamental and extensively studied topics in the database field. In this talk, I will introduce a certain class of database query that is not well supported by current indexing techniques. However, efficient handling of such queries are essential to a wide range of database applications, including those in system management, bioinformatics, XML information systems, etc. To serve these applications, I present the sequence data model as well as an index structure designed specifically for the data model. I will show how data in different forms, including sequential, relational, and semi-structured, can be mapped nicely into this data model, and how queries over those different data are efficiently supported by the index structure.

Seminars 2: Clustering, Classification, and other Data Mining Techniques

Time: 8:30-11:30am, Dec 4, 2003 Classroom: **4605#**

Data mining is a research area spanning several disciplines, including database systems, machine learning, statistics, etc. For instance, clustering and classification, which are among the core techniques of data mining, have been studied extensively in machine learning. Nevertheless, new applications, especially data-intensive applications, have constantly introduced practical and

theoretical challenges to these techniques. As our first example, I will show applications in E-commerce, bioinformatics that require new distance models to measure pattern similarity. This opens up many new research issues to clustering. As a second example, we study classification in the new environment of streaming data that has time-changing distributions. I will present and compare several different solutions to these challenges.

Seminars 3: Advanced Topics in Database Language and Systems

Time: 2:00-4:30pm, Dec 4, 2003 Classroom: 信息楼配楼 2 层多媒体教室

A lack of power and extensibility in their query languages has seriously limited the generality of DBMSs and hampered their ability to support new applications domains. Considerable efforts by database researchers and commercial DBMS vendors were devoted at overcoming these limitations, and led to the introduction in commercial DBMSs of significant extensions, including those for ROLAPs, recursive queries, and datablades. However, the effectiveness of these extensions is restricted to specific application domains, and there remain important applications are not supported well in SQL3. Thus, there is a pressing need for more general mechanisms for extending SQL and dealing with new application areas. In this talk, I will present the ATLaS system, which answers this need by means of generalized table functions and aggregate functions defined using the SQL language itself, rather than an external programming language. This simple extension turns SQL into a powerful database language, which can express a wide range of applications, including recursive queries, ROLAP aggregates, time series queries, and data mining functions. In addition to bringing great power and flexibility to SQL, the ATLaS system achieves excellent performance..

Biography

Haixun Wang is currently a research staff member at IBM T. J. Watson Research Center. He received the B.S. and the M.S. degree, both in computer science, from Shanghai Jiao Tong University in 1994 and 1996. He received the Ph.D. degree in computer science from the University of California, Los Angeles in 2000. He has published more than 40 research papers in referred international journals and conference proceedings (VLDB, SIGMOD, ICDE, etc.). He is a member of the ACM, the ACM SIGMOD, the ACM SIGKDD, and the IEEE Computer Society. He has served in program committees of international conferences and workshops, and has been a reviewer for some leading academic journals in the database field.

有意参加者请回 email 确认，并请宣传。

联系人 中国人民大学信息学院 王宇

Email: sandpiper@263.net

Tel: 6251 5575