

Xiaoyu Wang

wangxy@cs.brandeis.edu
(617)515-6033

70 Hall Street Apt. 4
Waltham, MA 02453

Education

- Ph.D., Computer Science* December, 2006
Topic: Reducing the Costs of I/O Intensive Operations during Query Processing
Advisor: professor Mitch Cherniack
Brandeis University, Waltham, MA.
- M.S., Computer Science* December, 2000
Brandeis University, Waltham, MA.
- B.Eng., Computer Science and Technology* July, 1999
Tsinghua University, Beijing, China

Skills

Systems: Linux, Windows, UNIX
Languages and tools: C/C++, Java, SQL, ML, Perl, Latex
Software: Postgres, SQL Server, DB2, etc.

Research Experience

Graduate Research Assistant, with professor Mitch Cherniack Fall 1999 – Present

- *Permuting and refining block request sequences.*
I/O intensive operations, e.g. sort and hash join, store intermediate results on disk temporarily when input relations are too large to be processed in memory. They send read and write system calls (block requests) to the underlying system (database system or operating system), which determines how to satisfy these requests. Because the underlying system uses strategies like caching, I/O merging to optimize its own performance, block requests are processed at different costs when appearing in different orders. Permuting and refining block request sequences is about changing the order of block requests produced by I/O intensive operations so that their block requests can be processed efficiently by the underlying system.
- *Avoiding unnecessary sorting and grouping operations during query processing.*
Sorting and grouping are two of the most expensive operations during query processing. Decision support queries generally involve several sorting and grouping operations. By formalizing the representation of ordering, grouping, and sub-ordering properties and inferring those properties through a query plan, unnecessary sorting and grouping operations in query plans can be identified and therefore eliminated. This technique is implemented in *Postgres* with C.

Teaching Experience

TA, *Introduction to Database Systems* (CS127b) Fall 2000 and Fall 2002

- This course teaches database designing, querying, tuning, etc.
- TAs are in charge of designing course projects, grading home works/projects, holding tutorial sessions, and teaching some chapters.
- The projects are done with *Postgres* database systems. The languages used are SQL, PG/PLSQL, JDBC, and Java.

TA, *Computer Structure and Organization* (CS31a) Fall 1999 and Fall 2001

- This course teaches operating systems, including process scheduling, memory management, file systems, etc.
- The course projects use Java on Linux.

TA, *Topics in Systems: Querying the Web* (CS120a) Spring 2000

- This course introduces new topics in web-related database researches. It involves reading, presenting, and discussing research papers.

Publications

Xiaoyu Wang and Mitch Cherniack. "Improving Query I/O Performance by Permuting and Refining Block Request Sequences" CIKM 2006.

Xiaoyu Wang and Mitch Cherniack. "Avoiding Sorting and Grouping in Processing Queries." VLDB 2003: 826-837.

Wenyin Liu, Xiaoyu Wang, Long Tang and Dov Dori. "Impact of Sparse Pixel Vectorization Algorithm Parameters on Line Segmentation Performance." GREC 1999: 335-344.

Honors and Awards

Full time graduate teaching/research assistantship from Brandeis September, 1999 – Present

Outstanding student awards of Tsinghua university 1995 – 1998

Tsinghua-Holison Scholarship for outstanding student December, 1994

Miscellaneous

Student member of ACM and New England Database Society

References

Professor Mitch Cherniack (Brandeis University)

Professor Stan Zdonik (Brown University)

Professor Timothy J. Hickey (Brandeis University)