Navin Kabra

101B, Twin Towers, D.P. Road, Aundh, Pune, 411 007. Phone: +91 98220 20096
E-mail: navin (at) smriti.com. WWW: http://www.cs.wisc.edu/~navin
Linked-in: http://www.linkedin.com/in/navinkabra

For the latest copy of this resume see: http://smriti.com/resume/. *At this time, I am not interested in a regular job - please do not contact me with full-time job offers.*

Areas of Interest

Content management, classification, and search; Growing and nurturing websites and online communities; Distributed and fault-tolerant software Systems; File-systems and storage technology; Design and implementation of database systems.

Education

Ph.D. in Computer Sciences, University of Wisconsin–Madison, USA, June 1999.
Advisor: Prof. David J. DeWitt
Dissertation: *Query Optimization for Object-Relational Database Systems*

M.S. in Computer Sciences, University of Wisconsin–Madison, USA, May 1994.

B.Tech. in Computer Science, Indian Institute of Technology-Mumbai, India, May 1992.

Work Experience

Currently on a sabbatical, Pune, India, Dec 2007–*present*
I am currently on a sabbatical – I am taking time off to explore some ideas I have about information processing on the internet, and to generally get a better understanding of how to build online communities organized around specific areas of interest.

I am also doing some small consultation projects in my areas of interest.

Symantec Corporation (formerly Veritas), Pune, India, Aug 2002–Dec 2007
Roles: Senior Researcher; CTO's Staff
I was a part of *Symantec Research Labs (SRL)*, which builds prototypes of emerging technologies to determine whether they can be productized. Some of the projects I was involved in: 1) the use of statistical techniques to analyze corporate communications data (*i.e. e-mail*) for preventing leakage of sensitive data, 2) the use of data-mining algorithms to automatically detect configuration anomalies and other mistakes in large enterprise data-centers, and 3) the application of information retreival algorithms to automatically detect variants in malware samples. In some projects, I had a hands on role where I did everything from the conceptualization and design, to the actual implementation and programming. In other projects, my role was more architectural - setting the direction, guiding the team, resolving conflicts, and evangelizing the idea across the company. I also worked in Symantec India's CTO office on various programs to increase technical vitality in the company, to foster innovation, to improve the patents programme, and to mentor and guide junior engineers.

Quiq Incorporated, Madison, Wisconsin, USA and Pune, India, Feb 2000–Jun 2002
Role: Senior Product Architect and Principal Developer
Quiq Inc was a company that provided software and services for Internet based Customer Support and eCRM (Customer Relationship Management). I was responsible for the design and implementation of the core search engine used in the product suite. We devised a novel index structure that incorporated both relational as well as text data. It provided for dynamic updatability (to reduce scheduled downtime), partitioned and replicated parallelism (for high availability), recoverability (from crashes as well as media failure), and support for a large number of simultaneous users.

Teradata Corporation (NCR Corp. at that time), Madison, Wisconsin, USA, Mar 1998–Jan 2000
Role: Module Architect and Project Leader
Worked on the Teradata Object-Relational DBMS (TOR), a scalable, parallel, object-relational

database management system. I was involved in various aspects of the architecture of the system, and design and development of individual features. Specific work includes: sole responsibility for design and implementation of user-defined functions (UDFs) in TOR, and of views in the query language; project leader of 2-person and 3-person teams to design and implement generalized user-definable aggregate operators in TOR, and design of text and spatial "data-blades" based on third-party software.

University of Wisconsin–Madison, Wisconsin, USA, Jun 1993–Feb 1998
Role: Research Assistant
Part of the team that designed and developed the Paradise Scalable Object-Relational DBMS. I worked on the project from its inception until it was acquired by Teradata Corporation (part of NCR at that time) in 1998. I had full responsibility of a number of modules of the software including the query parser, the optimizer, and parts of the scheduler. In addition, I worked on various aspects of the system including the extended data-types, the client interfaces, and the system catalogs.

University of Wisconsin–Madison, Wisconsin, USA, Sept 1992–May 1993
Role: Teaching Assistant
Conducted lectures and discussion sessions on Numerical Methods, Data Structures, and Introductory Pascal.

Indian Institute of Technology–Mumbai , India, Sept 1990–Dec 1991
Role: Teaching Assistant
Conducted discussion sessions for an introductory computing course for college freshmen.

Ph.D. Thesis Research
Dynamic Query Optimization in Database Query Processing:
Designed and implemented *Dynamic Re-Optimization*, an algorithm that dynamically detects sub-optimality of a query execution plan during query execution and improves performance by re-optimizing the query. Statistics are collected at key points during the execution of a complex query, and are then used to optimize the execution of the query, either by improving the resource allocation for that query, or by changing the execution plan for the remainder of the query. To ensure that this does not significantly slow down the normal execution of a query, the Query Optimizer carefully chooses what statistics to collect, when to collect them, and the circumstances under which to re-optimize the query.

Extensible Query Optimization:
Designed and Implemented OPT++, a tool that uses an object-oriented design to simplify the task of implementing, extending, and modifying an optimizer. Building an optimizer using OPT++ makes it easy to extend the query algebra (to incorporate new query algebra operators and physical implementation algorithms in the optimizer), easy to change the search space explored, and also easy to change the search strategy used. Furthermore, OPT++ comes equipped with a number of optimization techniques and search strategies that are available for use by an Optimizer-Implementor. Conducted a performance study that validates the design of OPT++ and shows that in spite of its flexibility, OPT++ can be used to build efficient optimizers.

Query Optimization in Object-Relational Database Systems:
Used OPT++ to implement and study a number of different optimization techniques and search strategies and how they interact with each other. Implemented a number of search strategies including dynamic-programming (System-R style), Simulated Annealing, Iterated Improvement, Two-Phase Optimization, and A*. Implemented a number of optimization techniques to handle join enumeration, expensive predicates, reference-valued attributes, path indexes, set-valued attributes, abstract data-types with methods, and spatial operations. For each optimization technique, studied how effective it is, and how it is affected by the choice of search strategy.

Refereed Publications
"Mass Collaboration: A Case Study" with Raghu Ramakrishnan *et. al. International Database*

Engineering and Applications Symposium (IDEAS-04). Coimbra, Portugal, July 2004.

"The QUIQ Engine: A Hybrid IR–DB System." with Raghu Ramakrishnan and Vuk Ercegovac. *International Conference on Data Engineering*, Bangalore, India, March 2003.

"Opt++: An Object-Oriented Design for Extensible Database Query Optimization." with David J. DeWitt. *The VLDB Journal*. Volume 8 Issue 1, January 1999.

"Efficient Re-Optimization of Sub-Optimal Query Execution Plans." with David J. DeWitt. *Proceedings of the 1998 SIGMOD Conference*, Seattle, Washington, June 1999.

"Building A Scalable GeoSpatial Database System: Technology, Implementation and Evaluation." with the Paradise Team. *Proceedings of the 1997 SIGMOD Conference*, Tucson, Arizona, May 1997.

"Client-Server Paradise." with David J. DeWitt, Jun Luo, Jignesh M. Patel and Jie-Bing Yu. *Proceedings of the 20th VLDB Conference*, Santiago, Chile, September 1994.

Selected Patents
"Unified Database and Text Retrieval System." with Raghu Ramakrishnan, Uri Shaft and Vuk Ercegovac. *U.S. Patent Number 6681222.* Assigned to Quiq Incorporated.

"Method and apparatus for evaluating index predicates on complex data types using virtual indexed streams." with Jignesh Patel. *U.S. Patent Number 6,678,686.* Assigned to NCR Corporation.

"Method and apparatus for using Java as a stored procedure language and as an embedded language on a client." with Ashutosh Singh and Jignesh Patel. *U.S. Patent Number 6,477,540.* Assigned to NCR Corporation.

"Method and apparatus for parallel execution of SQL-from within user defined functions." with Jignesh Patel, Jie-Bing Yu, Biswadeep Nag and Jian-Jun Chen. *U.S. Patent Number 6,594,651.* Assigned to NCR Corporation.

See also U.S. Patent Numbers 6,732,084 and 6,507,834. In addition, my work with Symantec Research Labs has resulted in over 14 patent applications that are currently pending.

Other Activities
Invited Talk, Resilient Computing
Given at the *National Conference on e-Security and Business Process Continuity*, organized by Maharastra Association for Information Technology (MAIT) and Indian Banks Association (IBA) at Mumbai, India, July 2004.

Invited Talk, Peer-to-Peer: The Future of Storage?
Given at the *Storage Networking Summit*, Mumbai, India, December 2003.

Managing Committee Member, Computer Society of India, Pune Chapter, June 2003–May 2004.

Program Committee Chair, CSI-Pune Workshop on Management of Large Data-Sets, December 2003

Websites and online communities, 1995–present
`http://punetech.com/`
`http://designforpune.com/`
`http://smriti.com/`
`http://withoutgivingthemovieaway.com`
`http://smritiweb.com/navin`
In my spare time, I have created and maintained various websites and blogs which together get over 10,000 visitors per day. This has helped me get a very good understanding of how the web works, and of growing and nurturing websites and online communities. This understanding has been

very helpful for me in my role on the CTO's Staff in Symantec, and I believe will be increasingly important and as "web-2.0" technologies and processes get adopted across the board.

References available upon request.