Naveen Koorakula
naveen@cs.unc.edu
http://www.cs.unc.edu/~naveen

57 Deming St,
San Francisco CA, 94114
Phone: 415 425 5412

Profile Summary Stellar track record in recognizing opportunities, initiating projects, developing and **driving core search technology** systems. Experience being a **thought leader in the cluster management space**, at companies with large deployments. Led and contributed individually to **major re- architecture projects at Yahoo Web Search**, which directly resulted in significant reductions in build times, with direct impact on key metrics like freshness and result quality. Proven ability to **manage large scale projects spanning multi- functional diverse teams**. **Self- starter** who recognizes and initiates efforts to help organization in any manner possible.

Areas of Interest Clusters, Grids / Utility Computing, Distributed Systems, Collaborative Systems, Operating Systems, Algorithms, Real Time Scheduling

Education **University of North Carolina at Chapel Hill:**
• Master of Science, Computer Science, August 1998 – May 2000
Indian Institute of Technology, Madras:
• Bachelor of Technology, Computer Science and Engineering, July 1994 - May 1998

Work Experience **Senior Software Engineer, Technical Lead, Inktomi/Yahoo Web Search** (Aug 2002 – Dec 2005)
• **Proposed, designed and led a** major re- architecture of the Echelon Cluster management system
  ○ Echelon is the **core cluster management and automation system** responsible for building web indexes and keeping them fresh, at Yahoo.
  ○ Designed and implemented
    • centralized job management and storage management
    • support for heterogeneous hardware
    • an *efficient scheduler* with
      ○ undesirable conditions in the cluster defined as *exceptions* (Eg:, slice of database not built)
      ○ all possible strategies defined as *handlers* to resolve exceptions, each with associated resource requirements and expected completion times
      ○ the scheduler dynamically chooses the best handler to handle exceptions in the system, given the current resource usage profiles of each node.
    • capability for echelon nodes to handle multiple jobs concurrently
    • support for automatic in- cluster mirroring of databases
    • support for using databases from similar clusters to populate the index faster
  ○ As a direct result of this re- architecture, Echelon clusters became much more robust, with the **index build turn- around shortening significantly**. Among other things, Echelon clusters
    • *survived a major data center move*, without losing data or work done to that point
    • *survived catastrophic failures* like power failure on half the cluster, serving data as usual

• Managed the Yahoo web search index build and deployment process over a period of nine months
  ○ **Managed the entire process from end to end**. Including defining build objectives, acceptance criteria, resource allocation, metrics definitions (such as web coverage, relevancy thresholds etc), relevancy testing, data center deployment and post- deployment monitoring.
  ○ This is resource intensive role which involved cross functional coordination between virtually all the teams in Yahoo web search.
  ○ While in this role, initiated and co- ordinated an effort to formalize and codify the index build process – including well- defined roles for all the stakeholders and interfaces between them.

• Proposed the design for and **led a full re- architecture and rewrite of the Yahoo Web Search Indexer**
  ○ Introduced a semantic document object model, with
    • Document features, like words, tags etc, mapped to lexical *scopes* in the doc content
    • *Semantic tokens* associated with *scopes*
    • *Semantic views* defined as collections of tokens
    • Simple but powerful interfaces to operate on multiple views at once, to enable more complex feature identification and processing
    • An additive model of document content analysis, with no information loss
    • Designed and implemented the document object library from scratch, with these features
  ○ Document object model deployed in all of Inktomi/Yahoo Content systems

- Introduced a new XML serial representation for documents, for use in content systems
  - Made this the exchange format of choice for passing documents between systems
  - This enabled validation at interfaces (where it wasn't possible before)
- This re- architecture has resulted in **enabling numerous innovations in document analysis** at Yahoo, which were not possible (or were extremely hard to implement) before.

- **Major contributor to the <u>Inktomi/Yahoo Page Cache</u>** design and implementation
  - Designed and implemented the page cache database format, associated tools and libraries
  - Designed and implemented support for caching web pages in the Content system
  - Designed and guided the development of the Page Cache display engine back- end
  - This contribution <u>closed a major hole in the Inktomi/Yahoo Web search product</u> offering.

- **Several other key contributions**
  - Designed and implemented an index data- correctness verification architecture. This is used as an automated system to make sure the search index has no major data flaws.
  - Designed and implemented the first version of a smart document abstract generator (snippets)
  - Acted as an evangelist for automated testing frameworks as an important cog in the development organization

- <u>Started a bi- weekly paper reading group</u> to discuss interesting recent publications and ran it for two years. This was mentioned by several people as an important and integral component of their on- the- job learning experience.

Software Engineer, Content Networking Services Group, Inktomi (Oct 2000 to Aug 2002)
- Contributed to the design and implementation of the Broadcast Manager suite of tools, a wide area network monitoring system for Media Distribution Networks (MDN)
- Individual contributor for the implementation of the Service Manager *service* architecture, used to effectively manage MDN deployments, with order of hundreds of nodes
- Individual contributor to the windows- media load simulator, which involved reverse engineering parts of the windows media protocol

Software Engineer, Fastforward Networks (June 2000 – Oct 2000)
- Individual contributor in a team developing the launcher, a tool to configure and launch services on every node of an MDN

Summer Intern, Exchange group, Microsoft (May 1999 - Aug 1999)
- Implemented support for OLEDB and Local Store services on various Windows platforms

Graduate Research Assistant, Dept. of Computer Science, UNC- CH
- Collaborative Systems – Jan 2000 to May 2000.
- TUNE research group – Dec 1998 to May 1999.

Graduate Teaching Assistant, Dept. of Computer Science, UNC- CH
- COMP 202 Algorithms – Aug 1998 to Nov 1998 & Aug 1999 to Dec 1999.

Teaching Assistant, Indian Institute of Technology, Madras,
- CS110 – Introduction to Computing - freshman class, July 1997 to Dec 1997.

Publications   V. Roussev, P. Dewan, N. Koorakula and S. Sellappa, *Integrating XML and Object- Based Programming for Distributed Collaboration*, Proceedings of the IEEE 9th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WET ICE '00)

School Projects

Real Time Systems
- Design of an improved Hierarchical Packet Pair Queuing (H- PFQ) algorithm for network traffic scheduling
- Design of a linear algorithm for scheduling periodic tasks for asymptotically optimal schedules

Distributed and Collaborative Systems
- Design and implementation of a Java Beans based collaborative editing tool for XML documents
- E- Advisor: a distributed collaborative system to automate faculty- student advising
  - http://www.cs.unc.edu/~stotts/COMP145/homes/advising

Operating Systems and Networking
- Design and implementation of a XINU-like distributed kernel with process management
  - Distributed IPC
  - Distributed terminal driver
- Implemented a basic UNIX shell with support for pipes and environment variables
- Implemented a Semantic File System using shell scripts
- Implemented a simple FTP protocol over DOS and UNIX systems

Algorithms
- Implemented Bottom-up sampled heaps - a novel, practical, worst-case efficient, probabilistic data structure for priority queues
- Implemented the Delaunay triangulation method for mesh generation

Compilers
- Compiler for a subset of the Pascal language using lex, yacc etc

Professional Skills
- *Languages*: C, C++, Tcl, Java, Perl, Intel 8086 and M68k assembly
- *Operating Environments*: UNIX (Linux, Solaris, *BSD,…), Windows (95, 98, NT, 2000), X
- *Other*: XML, HTML, TCP/IP, SNMP API

Honors
- Ranked 1 among STs, All India IIT-JEE 1994 (about 100,000 candidates)
- Ranked 86th (0.1% - more than 100,000 candidates), Engineering, Medicine, and Agricultural Common Entrance Test (EAMCET) 1994.
- Recipient of the National Merit Scholarship awarded by Govt. of India
- Topper of class throughout high school, including graduating exam.

Relevant Courses
Advanced Distributed Systems, Advanced Operating Systems, Parallel and Distributed Systems, Real time systems, Scheduling in Real time Networks, Collaborative systems, High Performance Computing, Concurrent and Distributed Algorithms, Computer Networks, Compilers, Computer Organization, Computer Architecture, Computer Systems Design, Advanced Programming Languages, Design and Analysis of Algorithms, Recent Trends in Theoretical Computer Science

Activities
- Publicity Co-chair for Saagar, UNC - South Asian Awareness Organization, '98-'99.
- Institute Captain for Carroms, IIT Madras, 1997-98.
- School Pupils Leader, Secunderabad Public School, 1991-92.
- National Social Service member, IIT Madras, 1994-95.
- Basketball, Carroms, Tennis, Cricket and Chess. College-champion in Carroms for three years.

References
Available on Request.