Nilesh Bansal

http://queens.db.toronto.edu/~nilesh nilesh@cs.toronto.edu

Education

First year PhD candidate in Computer Science, University of Toronto — Advisor: Prof. Nick Koudas
Research Interests: *search, text analysis, web mining, data management*

Master of Science in Computer Science, University of Toronto — Jan 2007
Thesis: *BlogScope: Spatio-temporal Analysis of the Blogosphere* — CGPA 4.0/4.0

Bachelor of Technology in Computer Science, Indian Institute of Technology, Bombay — May 2005
Thesis: *Caching Highly Dynamic Content: Formal Framework + Performance*

Research Experience

BlogScope: Spatio-temporal Analysis of the Blogosphere — Feb 2006 - Present
Advisor: Prof. Nick Koudas, University of Toronto
We are working on developing techniques for analysis of high volume text streams (e.g., weblogs). The aim is to design algorithms to automatically identify interesting time periods, map geographical locations, suggest surprising keywords, discover correlated terms, and generate topic summaries to aid information discovery. We are also looking at sub-problems arising in this process, e.g., ranked list aggregation in presence of hierarchies, and time series clustering.

This project has a live prototype system at **www.blogscope.net**[1]. I am responsible for the development, setup and management of BlogScope.net, which

- crawls and indexes weblogs in realtime as they are updated, and provides an interactive search and analysis interface.
- currently indexes over 75 million articles, with 12000 new articles added every hour.
- has traffic around three hundred thousand page views every month.
- is written in Java and built using open source software, using MySQL for storage, Lucene for indexing, Dojo toolkit for AJAX and a modern Web 2.0 interface, and with Tomcat application servers running behind an Apache httpd forward proxy server.

Caching Highly Dynamic Content: Formal Framework + Performance — Aug 2004 - Apr 2005
Advisor: Prof. Krithi Ramamritham and Prof. D. Manjunath, IIT Bombay
Developed and evaluated techniques for disseminating dynamic data to a large number of users from remote servers. Main approach used was to construct models for the stochastic behavior of the data and calculate cache times to minimize the probability of error.

Summarizing Tree Structured XML Data Quantitatively — May - Nov 2004
Advisor: Prof. Laks Lakshmanan and Prof. Raymond Ng, University of British Columbia, Vancouver
Developed and implemented an algorithm for constructing a summary of an XML document to discover the structural aspect of its schema, and to use the summary for tasks like query result size estimation, structural compression and exploration.

Publications

• Nilesh Bansal, Fei Chiang, Nick Koudas, Frank Wm. Tompa, *Seeking Stable Clusters in the Blogosphere*, In Proc. of the 33^{rd} Intl. Conference on Very Large Data Bases, VLDB 2007, Vienna, Austria, 2007.

• Nilesh Bansal, Nick Koudas, *BlogScope: A System for Online Analysis of High Volume Text Streams*, In Proc. of the 33^{rd} Intl. Conference on Very Large Data Bases, VLDB 2007 (Demo), Vienna, Austria, 2007.

• Nilesh Bansal, Nick Koudas, *BlogScope: Spatio-temporal Analysis of the Blogosphere*, In Proc. of the 16^{th} Intl. Conference on World Wide Web, WWW 2007 (Poster), Banff, Canada, 2007.

[1] Patent pending. See `http://www.blogscope.net/about/` for a complete list of features and details.

• Nilesh Bansal, Nick Koudas, *Searching the Blogosphere*, Proc. of the 10^{th} Intl. Workshop on Web and Databases, WebDB 2007, (co-located with SIGMOD) Beijing, China, 2007.

• Nilesh Bansal, Sudipto Guha, Nick Koudas, *Ad-Hoc Aggregations of Ranked Lists in the Presence of Hierarchies*, under review.

Technical Reports

• *Caching Highly Dynamic Content: Formal Framework + Performance*, Advisor: Prof. Krithi Ramamritham and Prof. D. Manjunath, Senior Undergraduate Thesis Report, IIT Bombay, May 2005.

• *Deep Processing of Group-bys for XML Analytics*, With: Amit Chandel, Prof. Laks Lakshmanan, and Prof. Raymond Ng, Tech Report, University of British Columbia, Sept 2004.

• *Issues in Efficient Data Access on Web*, Advisor: Prof. Krithi Ramamritham, Junior Undergraduate Thesis Report, IIT Bombay, April 2004.

Work Experience

Teaching Assistant, Dept. of Computer Science, University of Toronto 2005 - Present (5 Terms)
Duties include delievering tutorials, conducting labs, marking assignments and holding office hours.

System Administrator, Database Lab, University of Toronto Dec 2005 - Present
Management of computing facilities including web and database servers, storage, backups, thin clients, software installations, and other hardware owned by the lab.

Research Intern, Dept. of Computer Science, University of British Columbia May - Aug 2004
Worked on research project ISCOPE, designed and developed efficient summarization algorithm for XML data.

System Administrator, Hostel 7, IIT Bombay Dec 2001 - May 2005
Management of the hostel network for its 300 residents along with facilities including www/ftp servers, squid proxy, nfs/nis/ldap servers, firewall, jabber, dns server and media streaming servers.

Summer Intern, Media Labs Asia, IIT Bombay May - Aug 2003
Worked on project *aAQUA* that aims at providing a multilingual communication framework for rural (mainly agricultural) community of India. This involved developing a question-answering forum that can be used in villages without internet connectivity.

Honors and Awards

- Ranked 3^{rd} all-India in the IIT Joint Entrance Exam 2001 (among 150,000 candidates).
- Awarded the Institute Scholarship for Academic Excellence (2001) by IIT, Bombay.
- Recipient of the prestigious NTS scholarship awarded by the Govt. of India, and secured 7^{th} rank in the state level exam of the same.
- Secured 5^{th} rank in State Science Talent Search Examination conducted by the Govt. of Rajasthan, India. Also secured 2^{nd} position in the state level science quiz organized by the same.
- Awarded the certificate of merit, National Top 1% for National Standard Examination in Physics, at Higher Secondary Level.

Selected Projects

Cricket Animation Oct - Dec 2005
Designed an animation showing a cricket match. The main attractions of the animation were the city of Toronto, the cricket stadium with many people, and lighting effects; everything modelled as complex 3D objects designed from scratch using OpenGL in C++. This project won the second prize in the Wooden Monkey Hall of Fame, Fall 2005 (http://www.cdf.toronto.edu/~karan/monkey/fall2005/assgn.htm).

Network Intrusion Detection using Stide Methodology Feb - Apr 2005
Designed an intelligent system using Stide-methodology to automatically detect possible events of network intrusion. Developed an utility in Perl to create dictionaries of normal and intrusive events, and use these dictionaries to detect network intrusions by monitoring the network activity.

Design and Implementation of a Compiler Jan - April 2004

Developed a compiler for C-like language and generated intermediate representation similar to mnemonic code. The compiler supported arrays, structures, conditional and unconditional loops as in C.

Relevant Courses

- *Databases* Introduction to Database Systems, Implementation Techniques for RDBMS, Data Mining and Data Warehousing, Advanced Topics in Datamining, Advanced Database Management Systems, Web Search and Mining
- *Theory* Data Structures and Algorithms, Design and Analysis of Algorithms, Theory of Computation, Topics in Metric Embeddings, Linear Programming and Combinatorial Optimization, Foundations of Cryptography
- *Systems* Operating Systems, Computer Networks, Language Processors and Compilers, Abstractions and Paradigms of Programming, Functional Programming, Logic Design
- *Maths* Discrete Mathematics, Probability and Stochastic Theory, Systems Modeling and Analysis, Linear Algebra, Differential Equations
- *Others* Topics in Software Engineering, Computer Graphics, Artificial Intelligence, Human Computer Interaction

Technical Skills

- *Languages* Java (JSP, JDBC, Servlet, Swing), C, C++ (Qt, KDE libs), Scheme
- *Scripts* Unix Shell, Perl/CGI, PHP, Python
- *Databases* RDBMS (Oracle, PostgreSQL, MySQL), XML (SAX, DOM), Lucene
- *Tools* Eclipse, CVS, SVN, Web (HTML, Javascript, AJAX)
- *Platforms* Linux, and other unix flavors

Extracurricular Activities

Organizational Activities July 2001 - May 2005

I have been an active participant in various activities, including volunteer in organization of Mood Indigo 2001 and Techfest 2002 (largest cultural and technology college festivals in India respectively), member of MLC 2004-2005 (student committee for management of university computing facilities), and student judge for Open Software 2005 (a national level software development competition). I was awarded many prizes by my hostel for the same, including Best Office Bearer 2002, Organizational Color 2003, Organizational Color 2005 and Graduating Citation 2005.

GNU Bharti, India May 2002 - May 2005

GNU Bharti is a non-profit organization working to promote use of computer based technology (using open software) in rural parts of India. I have worked with this organization as a volunteer. We have also organized various workshops for promoting open source technologies in different cities.

Open Source Development Ongoing

I actively participate in open source projects and interact with the community. I have contributed to projects including Informa, FFmpeg, Amarok, Xiph Theora, OpenFTP, and KDE FTPMonitor.

Nilesh Bansal
June 2007