

Roshan R Sumbaly

CONTACT INFORMATION	rsumbaly@gmail.com +1-415-640-4510	http://cs.stanford.edu/people/rsumbaly http://github.com/rsumbaly
INTERESTS	Building large-scale systems and data-mining	
EDUCATION	Stanford University, USA M.S. in Computer Science • Specialization - Database Systems • Teaching Assistant - Information Retrieval & Web Search, Data Mining	2008 - 2010
	BITS Pilani, India B.E. (Honors) in Computer Science • Cumulative GPA - 10.0/10.0 • Teaching Assistant - Computer Programming 1, Parallel Computing	2004 - 2008
EXPERIENCE	LinkedIn, USA <i>Software Engineer</i> (April 2010 - Present) • Working on Project Voldemort (http://project-voldemort.com) and Hadoop as a part of SNA (http://sna-projects.com)	
	Yahoo! Inc, USA <i>Technical Yahoo! Intern</i> (June 2009 - September 2009) • As a part of the Cloud Computing & Data Infrastructure team, incorporated various compression algorithms at three different tiers of the PNUTS / Sherpa distributed datastore, resulting in decrease of average round-trip latency.	
	Stanford University, USA <i>Graduate Research Assistant</i> (September 2008 - June 2009) • Worked in collaboration with the Computational Earth & Environmental Science group to port various sparse complex matrix solvers to NVIDIA GPU Clusters using CUDA.	
	Hewlett Packard Labs, India <i>Research Intern</i> (January - June 2008) • Proposed and built a prototype data integration middleware (based on Grid Monitoring Architecture), for aggregation of HP's enterprise data. Data integration was achieved using RDF & SPARQL.	
	Indian Institute of Science (IISc), India <i>Research Intern</i> (May - July 2007) • Worked in the Grid Applications Research Lab on prediction of job queue wait time in batch scheduled machines (like Torque, LFS) using historical logs. Proposed new metrics and algorithms while also building a generic simulator for replaying logs to test new clustering algorithms.	
	Bhabha Atomic Research Centre (BARC), India <i>Intern</i> (May - July 2006) • Worked on a scheduling algorithm, based on backfilling optimization and fairness policies, and deployed it on a 512 node cluster in the Supercomputing Research Facility at BARC. Also contributed to an inhouse distributed monitoring system.	
	International Conference on High Performance Computing (HiPC 2007) 'AIGA - Artificially Intelligent Grid Assistant' • Developed a Grid based Question Answering system capable of mining answers from distributed data-sets. • Published an article in IEEE Technical Committee on Scalable Computing (TCSC) Newsletter titled 'Deployment of a Natural Language Processing system on a Grid'	

PROJECTS

Stanford University, USA

'Update Summarization'

- Built a system which generates a summary of a multi-document dataset based on the assumption that the user has already read a given set of documents.

'Opinion Mining over Large News Datasets'

- Developed metrics and algorithms to determine the opinion about people by mining New York Times corpus (1.8 million articles spanning over 20 years)
- Implemented using Aster Data's nCluster - a Map-Reduce based RDBMS with infrastructure provided by Amazon EC2.

'Supervised Machine Learning Classifiers for Usenet newsgroup messages'

- Implemented variants of classical classifiers like Naive Bayes, SVM, Decision Trees and Nearest Neighbor methods. Analyzed various existing feature selection methodologies and proposed new domain specific features to enhance accuracy of the classifier.

'Encrypted Tweets'

- Built a client side symmetric-key encryption system for Twitter using Greasemonkey.
- Also build a proxy server capable of performing man-in-the-middle attack on SSL.

BITS Pilani, India

'Analysis and Implementation of Load Balancing Algorithms in Distributed Environments'

- Simulation of variants of the classical balls into bins load balancing algorithms using SimGRID Toolkit.

'Personalization using Link Analysis'

- Implemented various link analysis algorithms on browsing history for personalized recommendations.

SKILLS

- Programming: C/C++, Java, Python, SQL, OpenMP and MPI Parallel Programming
- Toolkits: Eclipse, Lucene, CUDA, Hadoop
- Platforms: Linux, Windows, Solaris, Mac OS X
- Worked on Amazon's EC2, S3, Elastic MapReduce and Google's App Engine. Managed AWS resources (~\$30K worth of computing time) for 50 students as TA at Stanford

ACHIEVEMENTS & AWARDS

- Recipient of CEES/RPSEA 2008-09 Fellowship for research in GPGPUs
- Recipient of UC Berkeley Fellowship & Purdue University Graduate Fellowship for 2008-09
- Recipient of Narotam Sekhsaria Scholarship for 2008-10
- Awarded BITS Pilani Alumni Global '30 under 30' Award in 2009
- Recipient of Dhirubhai Ambani Undergraduate Scholarship & BITS Merit Scholarship, for all four years of undergraduate studies
- Awarded the Gold Medal for highest GPA in 2004 batch of BITS Pilani
- Led a team of four to win the National Runners-up Prize at Microsoft's Imagine Cup 2007 for the project 'eduGRID'
- Founder & Student Coordinator, Linux User Group and CSD (Centre for Software Development) at BITS
- Coordinator, Conferences
 - 'Solaris and Open Solaris, Java : Now & Future, Web and Mobile Applications using NetBeans', University Days, Sun Microsystems
 - 'Microsoft Robotic Studio', Microsoft
 - 'Cluster & Grid Computing', CDAC : Centre for Development of Advanced Computing