C.V.Krishnakumar Iyer

Contact Information	739 E El Camino Real Apt 112 Sunnyvale, CA-94087	*Email:*cvkkumar@cs.stanford.edu Ph: 408-462-5745

Interests　　Large Scale Data Mining and Analysis, Machine Learning, Information Retrieval and Software Engineering

Education

Stanford University,　　　　　　　　　　　　　　　　　　　　　　　　*Sep '08 - Apr '10*
M.S. Computer Science　　　　　　　　　　　　　　　　　　　　　　　　GPA : 3.86 / 4.0
- Recipient of the prestigious *BPCL Scholarship* awarded by Bharat Petroleum Corporation Ltd. awarded to the elite few for graduate study in the US.
- *Relevant Coursework:* Machine Learning, Data Mining & E-Business, Elements of Statistical Learning, Computational Advertising, Information Retrieval & Web Search, Probabilistic Graphical Models, Data Mining, Transaction Processing and Distributed Databases, Network Analysis.

Birla Institute of Technology and Science, Pilani - Goa Campus　　　*Aug '04 - Jun '08*
M Sc.(Tech.) Information Systems　　　　　　　　　　　　　　　　　　CGPA : 9.95/10.0
- Ranked #1 in Information Systems Department.
- Recipient of the Silver Medal awarded by BITS to the second rank holder ampngst the graduating batch of 2008.
- *Relevant Coursework :* Artificial Intelligence, Data Structures & Algorithms, Software Engineering, Probability & Statistics, Discrete Structures for Computer Science, Data Mining.

Experience

Apple Inc.　　　　　　　　　　　　　　　　　　　　　　　　　　　*Apr '10 to present*

Software Engineer - Internet Services

Fraud Detection on AppStore
- Responsible for the end-to-end development of a fraud solution to reputation fraud on iTunes AppStore as a part of a small team; involved right from requirements gathering, and conception to design, implementation and evaluation.
- Worked on applying data mining, network analysis and machine learning techniques on massive scale data.
- Implemented a scalable version of the *Belief propagation* that works on k-partite graphs with billions of nodes.
- Responsible for communication of the working of system to different stakeholders.
- System currently in production.
- Used Pig, Java MapReduce, Mahout and HBase on the Hadoop ecosystem.

Account Creation Fraud Detection on iTunes
- Designed and implemented a high performance scalable system that used stream mining algorithm to risk creations on real time data streams.
- Implementated in Java.

Spam Detection on iTunes Ping
- Worked on spam detection on iTunes Ping using machine learning.
- Involved in feature engineering, implementation and evaluation of the system.

Intern - Internet Services　　　　　　　　　　　　　　　　　　　　　June '09 to Mar '10
- Worked on automated sentiment analysis and opinion mining from Micro-blogs using a combination of sophisticated machine learning and data mining techniques.
- Part of a two-member team that was responsible for the entire project from its conception and design to implementation and production deployment.
- Project involved use and analysis of several state-of-the-art machine learning techniques for feature engineering, feature selection, skew handling in datasets, model comparisons for supervised learning, ensemble techniques for classifiers and evaluation of results.
- Used Java, LibSVM, Weka.

Hewlett Packard Labs-India　　*Research Intern*　　　　　　　　*Jan '08 to June '08*
STAIR : System for Topical and Aggregated Information Retrieval.
- Developed the architecture and the prototype of STAIR - an IR system that applied of a combination of Collaborative analysis and Focused Crawling techniques on the web documents to provide personalized, consolidated information relevant to the user as an aggregated PDF document.
- Implemented in on Java using Lucene. The semantics were obtained from WordNet.
- Published as a HP Labs Technical Report in 2009.

Center for Study of Language and Information, Stanford University

Graduate Research Assistant   *Aug '08 to June '09*

Cognitive Assistant that Learns and Organizes (CALO)

- Member of a team working on CALO, a system that extracts decisions from multi-party meetings to enable the effective handling of feedbacks.
- Also involved in the evaluation of new features to decision extraction process.
- Implementation used Java and LibSVM.

Bhabha Atomic Research Center, Trombay *Project Intern at DRHR*   *May 06 to July 06*

Image Processing and Software Development for Simplifying Robot Trajectory Generation

- Built a system that extracts information of a continuous path from any arbitrary raw image using graph-theoretic methods and provides input to the indigenous *Sensor-cum-Manipulator*, a Parallel planar kinematic robot.
- Implemented using C, Matlab (for image processing) and VB (as a wrapper GUI).

Selected Projects

Recommendation Systems based on Delicious and Twitter

- Implemented a People Recommendation System on Twitter (in Python) by a combination of several algorithms, that included collaborative filtering and network analysis.
- Designed and implemented an URL Recommendation system by analysis of tags from *Delicious*.

Role Discovery in Social Networks using Dirichlet Multinomial Regression Based Topic Modeling

- Implemented an unsupervised algorithm for identification of hierarchical roles in an organization using a combination of Social Network Analysis, Spectral Clustering and a variant of LDA (Latent Dirichlet Allocation).
- *Instructor: Prof. Daphne Koller, Stanford University*

URL Recommendation Based on Asymmetric Tag Similarity and Diffusion-Based Grouping

- Implemented an URL Recommendation system by analysis of tag similarity using data from ShareThis and Delicious using MapReduce and Partition based Joins on top of the Aster Cluster.

Finding Answerers on Yahoo! Answers

- Designed and implemented a system for selection of most appropriate answerers in *Yahoo! Answers* based on textual, structural and other auxiliary information. The result could be used to determine the routing for new questions.

Analysis of Text Based Classifiers

- Implemented and analyzed the performance of different Naïve Bayes classifiers on the 20-Newsgroups dataset, using Java and Lucene.

Comparison of Similarity Search Algorithms over Inverted Indexes

- Implemented and analyzed the performance of commonly used indexing similarity search algorithms: Term-at-a-Time and Document-at-a-Time. Also optimized the algorithms with efficient index compression.

RefMed - A Physician Referral and Review Service

- Designed and developed a physician referral and review service that enables patients to review and rate the physicians and facilitates the physicians to recommend other doctors to their patients.

Time Table Generator

- Implemented a system for automating the process of timetable generation for a University with its constraints, using a sub-optimal graph coloring approach for Constraint Satisfaction.

Publications

C.V.Krishnakumar and Dr.Krishnan Ramanathan, *STAIR : A System for Topical and Aggregated Information Retrieval*, Proceedings of the International Conference on Intelligent Human Computer Interaction (IHCI) 2009.

R.T. Durai Prabhakaran, B.J.C. Babu, V.P. Agrawal, C.V. Krishna Kumar, *A knowledge-based system for constituent material selection in polymer composite product design* - Proceedings of ISRS-2006, International Symposium for Research Scholars, IIT-Madras.

Skills

Languages: Java, Pig, Python, C, SQL.
Frameworks: Hadoop (Cloudera Certified Hadoop Developer), Spring, HBase, Mahout
Tools and Platforms: Weka, LaTeX, Basics of R and Teradata
Operating Systems: Unix based Systems, Microsoft Windows.

Honors and Achievements

Consistent recipient of the *Merit Scholarship* awarded by BITS Pilani to the top 10 students across the batch.

Secured the First Prize at *OpenSoft* - the software construction contest conducted as a part of QUARK-07, the national level technological fest at BITS-Pilani, Goa Campus.

Recipient of the Merit Certificate, awarded to the top 0.1% of students, for proficiency in English in AISSCE from the CBSE, 2004.