

# CURRICULUM VITAE

**V.G.Vinod Vydiswaran**

6/4, "Sakalnagar", Baner Road, Pune - 411 007.

Phone: +91-20-2565-3402

Email id: vgvinodv@it.iitb.ac.in

vgvinodv@yahoo.com, vgvinodv@hotmail.com

Homepage: www.it.iitb.ac.in/~vgvinodv

## EDUCATION

- **Masters of Technology (Information Technology)**  
**Kanwal Rekhi School of Information Technology (KReSIT)**  
**Indian Institute of Technology Bombay**  
July 2002 - date

Semester	I	II	III	IV
Semester Performance Index	9.63	9.71	9.85	
Credits Accrued	35	48	39	
Year	2002-03		2003-04	
Cumulative Performance Index	9.71	9.66	9.72	
	<b>(DEPT. TOPPER)</b>			

- **Bachelor of Engineering in Computer Engineering**  
**University of Pune**  
August 1998 - May 2002  
From Vishwakarma Institute of Technology, Pune - 37.

Semester	I	II	III	IV	V	VI	VII	VIII
% Marks	82.00	86.62	74.57	80.63	75.73	74.27	73.50	72.43
Class	Distn.	Distn.	Distn.	Distn.	Distn.	Distn.	Distn.	Distn.
Year	1998-99		1999-2000		2000-01		2001-02	
Year-wise aggregate	84.31 Distinction <b>(UNIV. TOPPER)</b>		77.80 Distinction <b>(UNIV. FIFTH)</b>		75.00 Distinction		73.00 Distinction	
Overall % (for 8 semesters)	<b>77.29 % (First Class with Distinction)</b>							

- **Higher Secondary School Education**  
**Central Board of Secondary Education (C.B.S.E.)**  
July 1996 - March 1998  
Kendriya Vidyalaya Ganeshkhind, Pune - 7.

86.80% (First Class with Distinction)

- **Secondary School Education**  
**Central Board of Secondary Education (C.B.S.E.)**  
July 1986 - March 1996  
Kendriya Vidyalaya Ganeshkhind, Pune - 7.

89.60% (First Class with Distinction)

## PERSONAL

- Date of Birth - February 27th, 1981
- Sex - Male
- Nationality - Indian
- Languages Known - English, Hindi, Tamil, Malayalam, Marathi, Sanskrit

## SKILL SET

- Data Mining & Information Extraction Techniques
- Algorithmic Analysis
- Statistical Modeling and Analysis
- Overall UNIX and LINUX Architecture and Tools
- Cryptography in Computer Systems
- Distributed File System Concepts

## EXPERIENCE

- **Teaching Assistantship** (July 2002 - date)
  - July - November 2002 : Distributed Systems
  - January - April 2003 : Data Mining and Warehousing
  - July 2003 - November 2003 : Implementation Techniques of Database Management Systems (in Distance Education mode)
  - January 2004 - date : Systems Laboratory
  - In addition, TA coordinator for the period July 2003 - date

## PUBLICATIONS

- Sunita Sarawagi, Sumana Srinivasan, V.G.Vinod Vydiswaran, Kapil Bhudhia. Resolving Citations in a Paper Repository. *In SIGKDD's newsletter "Explorations", December 2003 Issue.*

The paper describes the approach followed to tackle the Data Cleaning Task (Task 2) in KDD Cup 2003 organized in association with ACM's SIGKDD Conference, August 2003 at Washington D.C., USA. The entry bagged the second prize in the competition.

The paper is available at [www.it.iitb.ac.in/~vgvinodv/kdd2003.pdf](http://www.it.iitb.ac.in/~vgvinodv/kdd2003.pdf)

## PROJECTS

1. **M.Tech.Project: Learning Paths for Information Extraction**  
(Apr 2003 - date)

**Abstract:** With more and more data being made available online, there is a growing need to have an intelligent system that will help users search the webpages for specific information. The information available on the World Wide Web is in human-understandable form. Humans find it easy to browse the Web and get to the desired pages by clicking on a series of hyperlinks. They use visual clues from the webpages they visit to choose among probable links that would lead to a relevant page. Automation of this search process is possible if machines can be made to identify and quantify such clues and use them to decide on the most suitable links.

**Project Description:** The approach followed to automate the search process is to concentrate on one of the aspects of information extraction in which if the initial seed page is given and the desired goal state is defined, the machine would be able to identify a sequence of webpages (path) from the seed page to the goal page(s) through the webgraph. The project analyzes this problem and proposes probabilistic models as possible solutions. Approaches based on reinforcement learning, Hidden Markov Models (HMMs), and Conditional Random Fields (CRFs) are being considered.

**Remarks:** The project is an individual project guided by Prof. Sunita Sarawagi. Currently, the work is being done on sequence segmentation using HMMs and CRFs.

2. **KDD Cup Competition: Data Cleaning Task**  
(Apr - Jul 2003)

**Abstract:** KDD Cup Competition is an annual international competition held in association with ACM's SIGKDD Conference. The participation in the competition was done for the Data Cleaning task.

**Project Description:** The goal of the task was to re-create the citation graph of about 35000 papers given in L<sup>A</sup>T<sub>E</sub>X format. The files were cleaned to extract information such as title, authors, keywords, and bibliographic entries. By extracting keywords from the citation entries, the cited papers were identified using indices and feature extraction methods.

**Personal Contribution:** Extracted various features from the keywords, and developed different classifiers and tuned them to improve the estimation accuracy. Wrote data-cleaning and filtering scripts for feature generation. Involved in deciding confidence-value thresholds for citation prediction.

**Remarks:** The entry in the KDD Cup Competition was adjudged second on the symmetric difference criterion and the best in the number of correctly predicted edges in the citation graph.

3. **M.Tech.Seminar: Information Extraction - A Survey on Tools and Techniques used**  
(Sep - Dec 2002)

**Abstract:** The speed and efficiency with which Information Extraction is possible is a key factor in the success and usability of Information Retrieval systems. This seminar surveyed various algorithms and tools, including probabilistic models like HMMs, that are used for Information Extraction from structured and semi-structured text.

4. **Sanchay - A Massively Scalable and Secure Distributed File System**  
(Jul 2001 - Mar 2002)

**Abstract:** Sanchay is a distributed file system (DFS) designed to span the globe. The file system is designed to provide persistent access to user information from anywhere around the globe. The main aim of the project is to make the DFS scale to such large levels with consistency and security comparable to existing DFSs.

**Project Description:** This distributed file system is mounted on an underlying set of untrusted servers leased out to Sanchay by server-owners. Data is protected by cryptographic and redundancy replication techniques. All data leaving the client is encrypted and replicated at certain places globally. For performance reasons, all synchronous file system activities are placed in the kernel. Asynchronous activities like propagation of updates and encryption are implemented in user space. Maximum advantage of the existing physical file system has been exploited by adopting a stackable file system approach. To facilitate message passing and file transfer between objects in Sanchay to be user-transparent, a location-independent routing scheme is designed. The routing and administering information is stored in soft state and can be repaired or reconstructed very easily. Conflict resolution is preferred over complex distributed locking schemes. In addition, Sanchay supports loose semantics as in other file systems like NFS, Coda and Intermezzo.

**Personal Contribution:** Designed and implemented the authentication and authorization of clients and servers in the system. Designed the identification of users, servers and files in the system, which is a key factor in the location mechanism. Designed and implemented the encryption and decryption of data and metadata across the network. Also contributed in implementing the file system calls in the kernel for file accesses.

## 5. Content Search Index for Linux (Dec 2000 - Jan 2001)

**Abstract:** Owing to the increase of storage capacity of hard disks on personal computer systems, a large amount of time is spent on searching for data. It is ironical that a grep or a Find utility in Linux takes more time to search for a keyword on local storage media than what Google takes on the Net. This utility aims at providing a fast and efficient method to search through a set of files for a particular keyword.

**Description:** An index is created on all the keywords extracted from relevant files in predefined formats (e.g. .txt, .c, .cpp, .html, etc.). The user can write plug-ins to support other file formats. This index is stored on one of the physical partitions of the disk. Updates of the index are carried out by grouping mini-indices that are created on-the-fly whenever a file is added or modified, and merging them with the master index when the system is not being used heavily.

**Personal Contribution:** Contributed in designing the system. Implemented the modules for extraction of keywords from various file formats like .txt, .html, .c, .h and .cpp depending on the file types and the client

choices. These modules were then incorporated into the system, along with a generic interface for future coverage of other file formats like .pdf. Designed the algorithms used for searching keywords and optimizing storage.

6. **Database Management System for On-Campus Recruitment**  
(Mar 2001)

**Abstract:** A database management system was designed and developed with Visual Basic as the front-end and MS Access as the back-end to manage the database of students interested in applying for On-Campus Interviews. This system was extensively used for Recruitment procedure by the Training and Placement Office, Vishwakarma Institute of Technology, Pune.

7. **Health and Mathematics**  
(July - August 1995)

**Abstract:** Mathematics has a very close relation with the thought process. There is a fine balance between mind and logical thinking. The thought process can be deeply understood with the approach one adopts is solving puzzles.

**Description:** This was a research oriented experimental project taken up by me in Std. XI under the guidance of Dr.(Mrs.)B. Kaur, Assistant Principal, Kendriya Vidyalaya Ganeshkhind. We designed and developed puzzles of varying difficulty levels and recorded them in an audio cassette. This was used as a tool to analyze the human thinking process. Audio methodologies were used to observe how people represented information on paper, the importance they gave to the data they received and the approach they followed to solve the puzzles. This was then analyzed to derive common patterns.

## ACADEMIC ACHIEVEMENTS

1. Won the SECOND PRIZE in Data Cleaning Task (Task 2) in KDD Cup Competition held in association with ACM's SIGKDD Conference, Washington D.C., August 2003.
2. Won the FIRST PRIZE in the Veritas Open Software Contest at TechFest 2001 organized at IIT Bombay amongst 66 software projects from all over India for the project "Content Search as a Utility for the Linux O.S."
3. Won the THIRD PRIZE in the Open Software Competition at TechFest 2002 organized at IIT Bombay among 37 software projects from all over India for a partial prototype implementation of the project "Sanchay - A Massively Scalable and Secure Distributed File System".
4. Stood FIRST in PUNE UNIVERSITY in First Year (Common) Engineering Examination among 8596 students appearing for the same in 1998 - 1999.
5. Stood FIFTH in PUNE UNIVERSITY in Second Year Computer Engineering Examination among 1500 students appearing for the same in 1999 - 2000.
6. Selected to appear for INDIAN NATIONAL MATHEMATICAL OLYMPIAD (I.N.M.O.) after getting selected in REGIONAL MATHEMATICAL OLYMPIAD (R.M.O.) (16th in the Maharashtra - Goa Region) in Std.XI.
7. Member of the Zonal team participating in the 23rd Jawaharlal Nehru Kendriya Vidyalaya Sangathan's Science Exhibition in August 1995 held in New Delhi in Std.X for the project on "Health and Mathematics".
8. "MAHARASHTRA TALENT SEARCH" Scholarship holder: 25th in the state in Std.IX. Awarded Scholarship for the same. Awarded Certificate of Merit and Consolation Prize for Std.VIII.
9. All India Talent Search Award Certificate Of Merit in Std.V.
10. Throughout FIRST in all classes at School.

## EXTRA - CURRICULAR ACHIEVEMENTS

1. Elected General Secretary of Information Technology Association (ITA), KReSIT, for 2003.
2. Member of “Core Committee” for organizing workshops and giving lectures on Data Mining and its Applications (Confluence 2003), Linux and other Operating Systems (Prabhat 2003 and Confluence 2002).
3. Delivered invited lecture on “Post graduation options in India” to budding engineers at Vivekananda Education Society’s Institute of Technology (VESIT), Chembur, Mumbai. (October 2003)
4. Among the 3 finalists short-listed for the M/S. DGP Hinoday sponsored Vishwakarma Institute of Technology’s Best Outgoing Student Award for 2001-2002.
5. Chief Coordinator of the “BE Project Exhibition” in the Computer Engineering department, VIT in March 2001 and Member of the organizing committee for the same in March 2000.
6. Active member of the “Core Committee” involved in organizing “On Campus Recruitment Procedure” in VIT, in June 2001.
7. Peer educator for “AIDS Awareness and Sex Education” for Sevadham trust, Pune. Delivered awareness lectures to school-going children of standards VIII, IX and X from August 1997 to August 2001.
8. Certificate from the Ministry of Parliamentary Affairs, Government of India for Fourth Prize in National Youth Parliament Competition in Std. X and a member of the winning team in Regional Level and 7th in National Level.
9. Royal Commonwealth Society, London’s Highly Commended essay certificate in Std.IX.
10. Certificate of Appreciation from Embassy of the Republic of Korea for an essay on the topic “Strengthening India- Korea relations in the 21st Century” in Std. XI.
11. Obtained more than 50 ranking certificates for various Academic and Extra-curricular activities organized by Kendriya Vidyalaya Ganeshkhind, Pune and various other National Organizations.