

Ambuj K. Singh

EDUCATION

Ph.D. in Computer Sciences (December 1989)
The University of Texas at Austin

M.S. in Computer Sciences (May 1984)
Iowa State University, Ames, Iowa

B.S. in Computer Science and Engineering (May 1982)
Indian Institute of Technology, Kharagpur, India

EXPERIENCE

7/00 – present

Professor

Dept. of Computer Science, University of California at Santa Barbara.

7/95 – 6/00

Associate Professor

Dept. of Computer Science, University of California at Santa Barbara.

7/89 – 6/95

Assistant Professor

Dept. of Computer Science, University of California at Santa Barbara.

8/86 – 9/89

Graduate research assistant

Dept. of Computer Sciences, University of Texas at Austin.

Summer 88

Summer intern

Bell Communications Research, Morristown, New Jersey.

1/85 – 8/86

Teaching assistant

Dept. of Computer Sciences, University of Texas at Austin.

8/82 – 5/84

Teaching assistant

Dept. of Computer Sciences, Iowa State University.

RESEARCH AREAS

- Bioinformatics
- Databases and Data Mining
- Distributed Systems

SELECTED PRIOR WORKS

- “Top-k Spatial Joins of Probabilistic Objects,” (with V. Ljosa), 24th International Conference on Data Engineering, April 2008. To appear.
- “Efficient Algorithms for Mining Significant Substructures in Graphs with Quality Guarantees,” (with H. He), IEEE International Conference on Data Mining, October 2007.
- “MIST: Distributed Indexing and Querying in Sensor Networks using Statistical Models,” (with A. Bhattacharya and A. Meka), 33rd International Conference on Very Large Data Bases, Vienna, Austria, September 2007.
- “APLA: Indexing Arbitrary Probability Distributions,” (with V. Ljosa), 23rd International Conference on Data Engineering (Refereed), April, 2007.
- “GraphRank: Statistical Modeling and Mining of Significant Subgraphs in the Feature Space,” (with H. He), IEEE International Conference on Data Mining (ICDM), Dec. 2006.
- “Probabilistic Segmentation and Analysis of Horizontal Cells,” (with V. Ljosa), IEEE International Conference on Data Mining (ICDM), Dec. 2006.
- “Integrating Multi-Attribute Similarity Networks for Robust Representation of the Protein Space”, (with O. Camoglu and T. Can), *Bioinformatics* 22(13):1585–1592, 2006.
- “Index Structures for Approximate Matching in Sequence Databases,” (with T. Kahveci), *Handbook on Computational Biology*, editor Srinivas Aluru, CRC Press, 2006.
- “Closure-tree: An Index Structure for Graph Queries,” (with H. He), 22nd International Conference on Data Engineering (ICDE), April 2006.
- “Distributed Spatial Clustering in Sensor Networks,” (with A. Meka) 10th International Conference on Extending Database Technology (EDBT), pp. 980–1000, March 2006.
- “Indexing Spatially Sensitive Distance Measures Using Multi-Resolution Lower Bounds,” (with V. Ljosa and A. Bhattacharya), 10th International Conference on Extending Database Technology (EDBT), pp. 865–883, March 2006.
- “ViVo: Visual vocabulary construction for mining biomedical images,” (with A. Bhattacharya, V. Ljosa, J. Pan, M.R. Verardo, H. Yang and C. Faloutsos), Fifth IEEE International Conference on Data Mining (ICDM), November 2005. Received a best student paper award.
- “Analysis of protein-protein interaction networks using random walks,” (with T. Can and O. Camoglu), 5th ACM SIGKDD Workshop on Data Mining in Bioinformatics, August 2005.
- “Decision Tree Based Information Integration for Automated Protein Classification,” (with O. Camoglu, T. Can, and Y.F. Wang), *Journal of Bioinformatics and Computational Biology*, 3(3):717–742, June 2005.
- “DIST: A Distributed Spatio-temporal Index Structure for Sensor Networks, (with A. Meka), 14th ACM Conference on Information and Knowledge Management (CIKM), Nov. 2005.
- “A Unified Framework for Monitoring Data Streams in Real Time,” (with A. Bulut), 21st International Conference on Data Engineering, pp. 44–55, April 2005.

- “Distributed Data Streams Indexing using Content-based Routing Paradigm,” (with A. Bulut and R. Vienberg) In Proceedings of 19th International Parallel and Distributed Processing Symposium, April 2005.
- “Scalable Index Structures for Biological Data,” in Jason Tsong-Li Wang, Mohammed Javeed Zaki, Hannu Toivonen, and Dennis Shasha (Eds.), *Data Mining in Bioinformatics*, pp. 275-296, Springer, 2005.
- “A Distributed Database for Biomolecular Images,” (with B.S. Manjunath and R.F. Murphy), *SIGMOD Record*, 33(2), pp. 65–71, 2004.
- “Automated Protein Classification Using Consensus Decision,” (with T. Can, O. Camoglu, and Y.F. Wang), *IEEE Computer Society Bioinformatics Conference*, pp. 224–235, 2004.
- “Speeding up whole-genome alignment by indexing frequency vectors,” (with T. Kahveci and V. Ljosa), *Bioinformatics*, 20(13), pp. 2122–2134, 2004.
- “Index-based Similarity Search for Protein Structure Databases,” (with O. Camoglu and T. Kahveci), *Journal of Bioinformatics and Computational Biology*, 2(1), pp. 99–126, 2004.
- “Increasing Concurrency in Databases using Program Analysis (with R. Vitenberg and K. Kvilekval), 18th European Conference on Object-Oriented Programming (ECOOP), 2004.
- “Optimizing Similarity Search for Arbitrary Length Time Series Queries,” (with T. Kahveci) *IEEE Transactions on Knowledge and Data Engineering*, 16(4), pp.418–433, 2004.
- “Querying and Mining Biological Databases, *OMICS: A Journal of Integrative Biology*, 7(1), pp. 7–8, 2003.
- “Deriving Phylogenetic Trees from the Similarity Analysis of Metabolic Pathways,” (with M. Heymans), 11th International Conference on Intelligent Systems for Molecular Biology (ISMB), pp. 138–146, July 2003.
- “Faster Similarity Search for Multimedia Data Via Query Transformations,” (with C. Lang), *International Journal of Image and Graphics*, 3(1), pp. 3–30, 2003.
- “An Adaptive Scalable Middleware for Distributed Indexing of Data Streams,” (with A. Bulut and R. Vitenberg), *International Workshop on Databases, Information Systems and Peer-to-Peer Computing*, September 2003.
- “SWAT: Hierarchical Stream Summarization in Large Networks,” (with A. Bulut), 19th International Conference on Data Engineering, March 2003, pp. 303–314. Received the Best Paper Award.
- “Joining Massive High-Dimensional Datasets,” (with T. Kahveci and C. Lang), 19th International Conference on Data Engineering, March 2003, pp. 265–276.
- “MAP: Searching Large Genome Databases,” (with T. Kahveci), *Pacific Symposium on Biocomputing (PSB)*, January 2003, pp. 303–314.
- “Accelerating High-dimensional Nearest Neighbor Queries,” (with C. Lang), 14th International Conference on Scientific and Statistical Database Management (SSDBM), July 2003, pp. 109–118.
- “Dynamic Query-Based Debugging of Object-oriented Programs,” (with R. Lencevicius and U. Hoelzle), *Automated Software Engineering*, 10(1), January 2003, pp. 39–74.

- “Similarity Searching for Multi-Attribute Sequences,” (with T. Kahveci and A. Gürel), 14th International Conference on Scientific and Statistical Database Management (SSDBM), July 2002, pp. 175–184.
- “Prefetching for Mobile Computers Using Shape Graphs,” (with K. Kvilekval), Sixth Workshop on Languages, Compilers, and Run-time Systems for Scalable Computers (LCR02), June 2002.
- “An Efficient Index Structure for String Databases,” (with T. Kahveci), 27th International Conference on Very Large Data Bases, Rome, Italy, September 2001.
- “Modeling High-Dimensional Index Structures using Sampling,” (with C. Lang), ACM SIGMOD Conference, Santa Barbara, May 2001.
- “Variable Length Queries for Time Series Databases,” (with T. Kahveci), 17th International Conference on Data Engineering, Heidelberg, Germany, April 2001.
- “Recovering Distributed Objects,” (with J. James), *Information Processing Letters* 77(2-4): 143-150, 2001.
- “Stable and Fault-Tolerant Resource Allocation,” (with G. Johnson), Principles of Distributed Computing, Portland, July 2000.
- “Can a Shape Analysis Work at Run-time?,” (with J. Bogda), Java Virtual Machine Research & Technology Symposium, Monterey, March 2001.
- “Design of the Kan Distributed Object System,” (with J. James), *Concurrency: Practice and Experience*, 12(8): 755-797, 2000.
- “Dimensionality-reduction for Similarity Searching in Dynamic Databases,” (with K.V. Ravi Kanth, D. Agrawal, and A. El Abbadi), *Journal of Computer Vision and Image Understanding*, vol. 75, no. 2, pp. 59–72, 1999.
- “Efficient Dynamic Range Searching using Data Replication,” (with K.V. Ravi Kanth), *Information Processing Letters*, vol. 68, no. 2, pp. 97–105, 1998.
- “Scalable Access Within the Context of Digital Libraries,” (with X. Chung, R. Dolin, M. Neary, S. Prabhakar, K. V. Ravi Kanth, D. Wu, D. Agrawal, A. El Abbadi, M. Freeston, T. Smith, J. Su), *International Journal of Digital Libraries*, 1998.
- “A Framework for Programming with Non-atomic Memories,” *Journal of Parallel and Distributed Computing*, vol. 26, pp. 211–224, 1995.
- “Dynamic Query Based Debugging,” (with R. Lencevicius and U. Hölzle), *13th European Conference on Object-Oriented Programming (ECOOP)*, Lisbon, Portugal, June 1999.
- “Browsing and Placement of Multiresolution Images on Parallel Disks,” (with S. Prabhakar, D. Agrawal, A. El Abbadi, and T. Smith), *5th Annual Workshop on I/O in Parallel and Distributed Systems (IOPADS 97)*, San Jose, CA, November 1997.
- “Query-Based Debugging of Object-Oriented Programs,” (with R. Lencevicius and U. Hölzle), *ACM SIGPLAN Conference on Object-Oriented Programming Systems Languages and Applications (OOPSLA)*, October 1997, pp. 304–317.
- “Compositional Proofs for Concurrent Objects,” (with J. James), *International Symposium: Compositionality - The Significant Difference*, Malente/Holstein, Germany, September 1997.
- “Efficient View Maintenance at Data Warehouses,” (with D. Agrawal, A. El Abbadi, and T. Yurek), *ACM SIGMOD Conference*, Tucson, Arizona, May 1997.

- “Efficient Fault Tolerant Algorithms for Distributed Resource Allocation,” (with M. Choy), *ACM Trans. Programming Languages and Systems*, vol. 17, no. 3, pp. 535–559, May 1995.
- “Mixed Consistency: A Model for Parallel Programming,” (with D. Agrawal, M. Choy, and H. Leong), *Proc. 13th ACM Symposium on the Principles of Distributed Computing*, pp. 101–110, Los Angeles, CA, August 1994.
- “The Elusive Atomic Register,” (with J. Anderson and M.G. Gouda), *Journal of the ACM*, 41(2), 1994.
- Adaptive Solutions to the Mutual Exclusion Problem,” (with M. Choy), *Distributed Computing*, vol. 8, no. 1, pp. 1–17, 1994.
- “Program Refinement in Fair Transition Systems,” *Acta Informatica*, vol. 30, no. 6, pp. 503–535, 1993.
- “Consistency and Orderability: Semantics-Based Correctness Criteria for Databases,” (with D. Agrawal and A. El Abbadi), *ACM Transactions on Database Systems*, vol. 18, no. 3, pp. 460–486, 1993.

STUDENTS

- V. Ljosa, PhD dissertation: “Managing Probabilistic Data: Toward Data-driven Biology,” 2007; at Broad Institute.
- A. Bhattacharya, PhD dissertation: “Analysis of Neighborhood Relationships in Biomedical Images,” 2007; at Indian Institute of Technology, Kanpur.
- A. Meka, PhD dissertation: “Distributed Algorithms for Mining Spatio-temporal Data in Sensor Networks,” 2007; at Oracle.
- H. He, PhD dissertation: “Querying and Mining Graph Databases,” 2007; at Google.
- O. Camoglu, PhD dissertation: “Discovering Functional Relationships among Proteins Using Computational Techniques,” June 2006; at Ask.com.
- T. Can, Post-doctoral scholar, 2005; currently Assistant Professor, Department of Computer Engineering, Middle Eastern Technical University, Turkey.
- A. Bulut, PhD dissertation: “A Flexible Data Mining Architecture for Monitoring Data Streams,” 2005; at Citrix.
- R. Vitenberg, Post-doctoral scholar, 2005; at IBM.
- T. Kahveci, PhD dissertation: “Indexing and Querying of Sequence Databases”, June 2004; currently Assistant Professor, Dept. of Computer Science, Univ. of Florida.
- K. Kvilekval, PhD dissertation: “Predicting program events for mobile programming”, December 2004; currently Research Scientist, Center of Bioimage Informatics, UCSB.
- C. Lang, PhD dissertation: “Improving the Access to Tomorrow’s Data Repositories through Prediction”, January 2003; at IBM Research.
- J. Bogda, PhD dissertation: “Program Analysis Alleviates Java Synchronization”, (co-advised with U. Hölzle), Summer 2001.
- J. James, PhD dissertation: “Reliable Distributed Objects: Reasoning, Analysis, and Implementation,” completed Fall 1999; currently Assistant Professor, Department of Computer Science, Utah State University.
- R. Lencevicius, PhD dissertation: “Query-based Debugging,” (co-advised with U. Hölzle), completed Summer 1999; at Nokia Research Center, Burlington, MA.

- K. V. Ravi Kanth, PhD dissertation: “Multi-dimensional Indexing,” completed Spring 1998; at Oracle NEDC, Nashua, New Hampshire.
- Manhoi Choy, PhD dissertation: “Distributed Resource Allocation,” completed August 1994.
- Swaroop Jagadish, MS project, 2007; at Yahoo.
- G. Benjamin, MS project, 2003.
- H. Sun, MS project, 2002; at Microsoft.
- M. Heymans, MS, 2002; at Google.
- F. Hu, MS thesis: “Design and Implementation of Fault Tolerance for the Kan System,” 2001; at HP.
- G. Johnson, MS thesis: “Stable and Fault-tolerant Object Allocation,” 2000; at Microsoft.
- S. Brydon, MS, August 1999; at Sun.
- S. Bhatia, MS thesis: “Distributed Garbage Collection,” 1999; at Cisco Corporation.
- S. Lee, MS thesis: “Concurrency Control to Support Guarded and Nested Atomic Actions with Replicated Objects in Kan,” 1998; at Oracle Corporation.
- J. Wang, MS thesis: “Thread Optimizations in Concurrent Object Oriented Languages,” 1998; at Ericsson Corporation.
- C. Jones, MS thesis: “Low Latency MPI,” 1996; at Compaq Corporation.

CONTRACTS & GRANTS (04-07)

- Co-principal Investigator, “ITR: Collaborative Research on Next-Generation Bio-Molecular Imaging and Information Discovery,” National Science Foundation, \$9,400,00, 2003-2008.
- Principal Investigator, “Scalable Querying and Mining of Graphs,” National Science Foundation, \$505,594, 2006-09.
- Principal Investigator, “Scalable Homology Search Tools,” National Science Foundation, \$474,957, 2002-06.
- Principal Investigator, “Digital Campus: Scalable Information Services on a Campus-wide Wireless Network”, CISE Research Infrastructure Award, National Science Foundation, \$1,056,302, 2000-06.
- Investigator, “IGERT: Graduate Training Program in Interactive Digital Multimedia,” National Science Foundation, \$2,909,596, 2003-08.

HONORS

Invited speaker, Workshop on Bioimage Informatics, 2008.
 Keynote speaker, First Workshop on Data Mining of Uncertain Data, co-located with International Conference on Data Mining, 2007.
 Invited Speaker, Florida Bioinformatics Workshop, 2005.
 Invited Speaker, IEEE CSB Bioimage Workshop, San Jose, August 2005.
 Best Paper Award, 19th International Conference on Data Engineering, Bangalore, India, March 2003.
 MCD Fellow, The University of Texas at Austin, Fall 1984 – Spring 1986.
 National Merit Scholar, The Government of India, January 1969 – May 1976.
 Member, Phi Kappa Phi (General Scholastic Honorary).
 Member, Beta Alpha Phi (International Students Honorary).

**PROFESSIONAL
ACTIVITIES**

Member ACM, IEEE, ISCB.

Program co-chair, Workshop on Multiscale Biological Imaging, Data Mining, and Informatics, Santa Barbara, CA 2006.

Vice Chair, Program Committee, Track on Distributed Data Management Track, International Conference on Distributed Computing Systems, 2004.

Co-Chair, Program Committee, Workshop on Bioinformatics, International Conference on Data Engineering, 2003.

Corporate Sponsors Chair, SIGMOD/PODS 2001 Conference.

Local Arrangements, 17th ACM Symposium on Principles of Distributed Computing, August 1997.

Local Arrangements, 3rd Workshop on Self-stabilization, August 1997.

Member, Program Committee, International Conference on Data Engineering, International Conference on Distributed Computing Systems, IEEE International Conference on Multimedia and Expo, 2002. VLDB Conference, International Conference on Parallel Processing, International Workshop on Formal Methods for Parallel Programming: Theory and Applications, ACM Symposium on Principles of Distributed Computing, International Workshop on Distributed Algorithms.