

# Sameer Singh

140 Governors Drive  
Amherst MA 01003

+1 (562) 726 4373

✉ sameer@cs.umass.edu

📁 cs.umass.edu/~sameer

## Research Interests

- Machine Learning
- Graphical Models
- Information Extraction
- Scalability
- Distributed and Parallel Inference
- Semi-Supervised Learning

## Education

- 2008–2013 (exp) **PhD (Computer Science)**, *University of Massachusetts*, Amherst, MA      GPA: **4.0**.  
thesis      Scaling MCMC Inference, Belief Propagation and Learning to Large-Scale Graphical Models  
committee      Andrew McCallum (chair), Carlos Guestrin, Ben Marlin, David Jensen, Michael Zink
- 2004–2007 **MS (Computer Science)**, *Vanderbilt University*, Nashville, TN      GPA: **4.0**, with 4 A+.
- 2000–2004 **BE (Electronics & Comm Engg)**, *NSIT (Delhi University)*, New Delhi, India.

## Awards and Honors

- 2011 **Travel Award**, Neural Information Processing Systems (NIPS), Granda, Spain.
- 2011 **Best Talk Award**, Machine Reading Project Phase 3 Kickoff, Seattle, WA.
- 2010–2011 **Yahoo! Key Scientific Challenges (KSC) Award**, in Machine Learning & Statistics.
- 2010–2011 **Accomplishments in Search & Mining Award**, UMass CS Dept and Yahoo!.
- 2009–2010 **Graduate School Fellowship Award**, for continuing students, Univ of Massachusetts.
- 2009 **Travel Award**, Neural Information Processing Systems (NIPS), Vancouver, Canada.
- 2004 **Best Undergraduate Project**, Department of Instrumentation & Control (NSIT).

## Experience

### Industrial

- June–Sept 2010 **Research Intern**, *Google Research*, Mountain View, CA.  
Large-scale coreference using Map-Reduce based distributed inference & a novel hierarchical model
- July–Sept 2009 **Software Engineer Intern**, *Yahoo! Labs*, Santa Clara, CA.  
Unsupervised extraction of entities from Search Ads using constraint-based CRF learning
- May–Dec 2007 **Software Engineer Intern**, *Google Inc*, Pittsburgh, PA.  
Machine Learning and Feature Selection algorithms for Map-Reduce for the Ads Quality team
- May–Aug 2006 **Systems Engineer Intern**, *Adv Computing Center for Research & Education (ACCRES)*,  
*Vanderbilt University*, Nashville, TN.  
Implemented Schedulers for queries to a Distributed File System on Tape Storage
- Dec 03–Jan 2004 **Programming Intern**, *PRAVAK Cybernetics Ltd*, New Delhi, India.
- May–Aug 2003 **Robotics Intern**, *Central Electronic Engineering Research Inst (CEERI)*, New Delhi, India.  
Prosthetic hand that enables gripping of irregular, slippery and delicate objects

## Academic

- Jan 2008–present **Research Assistant**, *Information Extraction and Synthesis Lab (IESL)*, UMass, Amherst.  
Working on various Information Extraction tasks, applying Machine Learning to large scales
- Spring 2011 **Teaching Assistant**, *CS691: Probabilistic Graphical Models*, UMass, Amherst.
- Spring 2008 **Teaching Assistant**, *CS121: Introduction to Programming*, UMass, Amherst.
- May 05–May 07 **Research Assistant**, *Inst. for Software & Integrated Systems (ISIS)*, Vanderbilt University.
- Fall 04, Spring 05 **Teaching Assistant**, *CS201: Data Structures*, Vanderbilt University.
- 2003–2004 **Student Collaborator**, *Defense Research & Development Organisation (DRDO)*, India.
- 2002–2004 **Student Collaborator**, *School of Physical Sciences*, Jawaharlal Nehru Univ. (JNU), India.

---

## Professional Activities

- member Faculty Hiring Committee (Computational Biology), 2012
- organizer NIPS Workshop on Learning at Scale (Big Learning), 2011
- co-chair North-East Student Colloquium on Artificial Intelligence (NESCAI), 2010
- steward Computer Science, Graduate Employee Organization (GEO), UMass, 2010, 2011
- member New Students Committee, UMass Computer Science Department, 2009, 2010, 2012
- general secretary Vanderbilt India Association (VIA), 2005-2006
- reviewer AKBC 2010, NESCAI 2010, EMNLP 2010, KDD 2011, IJCAI 2011, EMNLP 2011, NIPS 2011, EACL 2012, OPTMAS 2012, AKBC 2012
- pc member NESCAI 2010, IJCAI 2011, EMNLP 2011, EACL 2012, OPTMAS 2012, AKBC 2012

## Invited Talks

- Dec 16 2011 **Parallel Approx BP**, *NIPS Workshop on Computational Tradeoffs for Statistical Learning*.
- Apr 6 2011 **Large-scale Cross-doc Coreference**, *Machine Reading Project Ph 3 Kickoff*, Seattle, WA.
- Dec 11 2010 **Distributed MAP Inference**, *NIPS Workshop on Learning on Cores, Clusters and Cloud*.
- Nov 5 2008 **Deep Belief Nets**, *Machine Learning and Friends Lunch (MLFL)*, UMass, Amherst.
- Nov 6,13 2006 **Symbolic Models of Cognition**, *Vanderbilt AI Research Group (VAIRG)*, Nashville, TN.

---

## Publications

### Book Chapters

- Jeremy Kubica, **Sameer Singh**, and Daria Sorokina. “Parallel Large-Scale Feature Selection”. In: *Scaling Up Machine Learning*. Cambridge University Press, 2011

### Refereed

- **Sameer Singh**, Amarnag Subramanya, Fernando Pereira, and Andrew McCallum. “Large-Scale Cross-Document Coreference Using Distributed Inference and Hierarchical Models”. In: *Association for Computational Linguistics: Human Language Technologies (ACL HLT)*. 2011
- **Sameer Singh** and Andrew McCallum. “Towards Asynchronous Distributed MCMC Inference for Large Graphical Models”. In: *Neural Information Processing Systems (NIPS), Big Learning Workshop on Algorithms, Systems, and Tools for Learning at Scale*. 2011
- **Sameer Singh**, Brian Martin, and Andrew McCallum. “Inducing Value Sparsity for Parallel Inference in Tree-shaped Models”. In: *Neural Information Processing Systems (NIPS), Workshop on Computational Trade-offs in Statistical Learning (COST)*. 2011
- **Sameer Singh**, Amarnag Subramanya, Fernando Pereira, and Andrew McCallum. “Distributed MAP Inference for Undirected Graphical Models”. In: *Neural Information Processing Systems (NIPS), Workshop on Learning on Cores, Clusters and Clouds (LCCC)*. 2010

- **Sameer Singh**, Limin Yao, Sebastian Riedel, and Andrew McCallum. “Constraint-Driven Rank-Based Learning for Information Extraction”. In: *North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT)*. 2010
- **Sameer Singh**, Dustin Hillard, and Chris Leggetter. “Minimally-Supervised Extraction of Entities from Text Advertisements”. In: *North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT)*. 2010
- **Sameer Singh**, Karl Schultz, and Andrew McCallum. “Bi-directional Joint Inference for Entity Resolution and Segmentation Using Imperatively-Defined Factor Graphs”. In: *Machine Learning and Knowledge Discovery in Databases (Lecture Notes in Computer Science) and European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*. 2009, pp. 414–429
- **Sameer Singh**, Jeremy Kubica, E. Scott Larsen, and Daria Sorokina. “Parallel Large Scale Feature Selection for Logistic Regression”. In: *SIAM Data Mining Conference (SDM)*. 2009, pp. 1171–1182
- Andrew McCallum, Karl Schultz, and **Sameer Singh**. “FACTORIE: Probabilistic Programming via Imperatively Defined Factor Graphs”. In: *Neural Information Processing Systems (NIPS)*. 2009
- Michael Wick, Khashayar Rohanimanesh, **Sameer Singh**, and Andrew McCallum. “Training Factor Graphs with Reinforcement Learning for Efficient MAP Inference”. In: *Neural Information Processing Systems (NIPS)*. 2009
- Andrew McCallum, Khashayar Rohanimanesh, Michael Wick, Karl Schultz, and **Sameer Singh**. “FACTORIE: Efficient Probabilistic Programming via Imperative Declarations of Structure, Inference and Learning”. In: *Neural Information Processing Systems (NIPS), Workshop on Probabilistic Programming*. 2008
- Tatiana Kichkaylo, Chris van Buskirk, **Sameer Singh**, Himanshu Neema, Michael Orosz, and Robert Neches. “Mixed-Initiative Planning for Space Exploration Missions”. In: *International Conference on Automated Planning and Scheduling Workshop (ICAPS)*. 2007
- S.R. Schach, T.O.S. Adeshiyam, D. Balasubramanian, G. Madl, E.P. Osses, **Sameer Singh**, K. Suwanmongkol, M. Xie, and D.G. Feitelson. “Common Coupling and Pointer Variables, with Application to a Linux Case Study”. In: *Software Quality Journal (SQJ)* (2007)
- D.G. Feitelson, T.O.S. Adeshiyam, D. Balasubramanian, Y. Etsion, G. Madl, E.P. Osses, **Sameer Singh**, K. Suwanmongkol, M. Xie, and S.R. Schach. “Fine-Grain Analysis of Common Coupling and its Application to a Linux Case Study”. In: *Journal of Systems and Software (JSS)* (2007)
- **Sameer Singh** and Julie A. Adams. “Transfer of Learning for Complex Domains: A Demonstration Using Multiple Robots”. In: *International Conference on Robotics & Automation (ICRA)*. 2006
- **Sameer Singh**. “Finding the shortest path for a mobile robot in an unmapped maze from minimum runs”. In: *International Conference on CAD, CAM, Robotics & Autonomous Factories (INCARF)*. 2003

### Unrefereed

- **Sameer Singh**, Michael L. Wick, and Andrew McCallum. “Distantly Labeling Data for Large Scale Cross-Document Coreference”. In: *Computing Research Repository (CoRR)* abs/1005.4298 (2010)
- **Sameer Singh**. “Option Discovery in Hierarchical Reinforcement Learning for Training Large Factor Graphs for Information Extraction”. In: *PhD Candidacy/Synthesis Report* (2009). Readers: Andrew McCallum, Andy Barto
- Khashayar Rohanimanesh, Michael Wick, **Sameer Singh**, and Andrew McCallum. *MAP inference in Large Factor Graphs with Reinforcement Learning*. Tech. rep. UM-CS-2008-040. University of Massachusetts, Amherst, 2008

### Patents

- **Sameer Singh**, E. Scott Larsen, Jeremy Kubica, and Andrew W. Moore. *Feature Selection For Large Scale Models*. US Patent Serial Number 12/263,008

---

## Relevant Courses

- Machine Learning
- Automated Knowledge Base Construction
- Fast and Parallel Processing (Math)
- Reinforcement Learning
- Computational Modeling (Psychology)
- Symbolic Architectures of Cognition
- Theory of Computation
- Neural Networks
- Advanced Artificial Intelligence
- Optimization in Machine Learning
- Distributed Artificial Intelligence
- Robotics
- Computational Cognitive Neuroscience
- Advanced Algorithm Design and Analysis
- Graph Algorithms

---

## References

Available upon request