Aditya Parameswaran
31 Angell Court #101
Stanford CA 94305

H: 650-497-2464 M: 650-353-1396
adityagp-AT-cs.stanford.edu
http://www.stanford.edu/∼adityagp

Research Interests

Algorithm design and analysis for databases, data mining and the internet.

Education

- **Stanford University** — **Stanford, CA**
  PhD Program, Computer Science (Recieved MS in Jun. 2010) — *Sep. 2007 - Present*
  - Recipient of the Terry Groswith School of Engineering Fellowship
  - Winner of the **Yahoo! Key Scientific Challenges Program** 2010
  - Passed the PhD Comprehensive Exams **in the first attempt** (the only student in my batch to do so)
  - Passed the Infolab PhD Qualifying Exams **in the first attempt**
- **Indian Institute of Technology Bombay** — **Mumbai, India**
  B.Tech. Computer Science and Engineering — *Aug. 2003 - Sep. 2007*
  - Graduated at the **top of my class** of 56 students in the CSE Dept, with a CGPA of 9.63/10.00
  - Awarded the **Institute Silver Medal** and the **Jayati Deshmukh Gold Medal** (2007)
  - Received **perfect GRE and TOEFL Scores** (2006)
  - Awarded the Aditya Birla Scholarship (2003) - A prestigious scholarship **covering the entire undergraduate education expenses** at IIT Bombay

Recent Research Experience

- **Optimizing Human Computation** — **Stanford University**
  Research Project: Guide: Profs. H. Garcia-Molina, Jennifer Widom, Alkis Polyzotis — *2009-*
  My thesis work involves leveraging people as processing units, much like computer subroutines, to achieve some global objective. Our focus is on *optimizing* this computation — while there may be many ways to orchestrate a particular task, our goal is to use as few resources (e.g., time, money) as possible, while getting equally good or better results as unoptimized computation.

  I am approaching this problem of orchestrating computing tasks involving people in two directions: (A) Building a system (DECO) that optimizes declarative queries over relational data and data crowdsourced on demand by minimizing the cost (of questions asked to humans), reducing uncertainty and improving performance. (B) Optimizing fundamental algorithms where the unit operations are performed by people. Examples of algorithms include: searching, filtering, sorting, clustering, classification, categorization, and so on. For more information, visit `i.stanford.edu/scoop`.
- **Crowdsourcing in Information Extraction** — **Yahoo! Research, Santa Clara**
  Research Project: Guide: Kedar Bellare, Nilesh Dalvi, Vibhor Rastogi — *Summer 2011*
  In ongoing work, I am working on the problem of using crowdsourcing (via human editors) to assist with the web information extraction pipeline in the PSOX group at Yahoo! Research.
- **Data Cleaning** — **Microsoft Research Redmond**
  Research Project: Guide: Raghav Kaushik, Arvind Arasu — *Summer 2010*
  I worked on a problem related to information extraction in the Data Cleaning project of the DMX group at MSR Redmond.
- **Generalized Recommendations** — **Stanford University**
  Research Project: Guide: Prof. Hector Garcia-Molina — *2008-10*
  I have been working on extending recommendation systems to work on domains where there are additional constraints to be satisfied. Specifically, we developed algorithms for recommendations for various prerequisite structures (and proved worst-case bounds), we examined the complexity of general requirement structures and developed recommendations that are not only "good" but also satisfy requirements and lastly, we developed probabilistic techniques for recommendations exploiting temporality in items being consumed. This ongoing research is in the context of the *CourseRank* project in the Infolab.

- **Information Extraction: Extending Wrappers** **Yahoo! Research, Bangalore, India**
  Research Internship: Guide: Rajeev Rastogi, Nilesh Dalvi *Winter 2009*
  At Yahoo! Research Bangalore, I worked on the problem of developing robust wrappers for information extraction from template based websites. We developed formal frameworks for robustness, and designed and experimented with optimally robust wrappers.

- **Synthesizing View Definitions from Data** **Stanford University**
  Research Project: Guide: Profs. Hector Garcia-Molina, Jennifer Widom *2009-10*
  I worked on the problem of finding the most succinct and accurate view definition given a database instance and a view instance when the query is restricted to a specific family of queries. We studied the tradeoffs between approximation, succinctness and family of queries. We are now looking to extend this work to assisted query synthesis.

- **Concept and Relationship Mining** **Kosmix Corp., Mountain View**
  Research Internship: Guide: Anand Rajaraman *Summer 2008*
  I worked on the problem of enriching Kosmix's concept taxonomy by discovering new concepts, and attaching them to suitable positions in the taxonomy. We developed algorithms to automatically evaluate and extract concepts from a web corpus using statistical techniques. We also designed techniques to determine relationships between extracted concepts. These techniques were implemented for a dataset of query logs and web documents.

- **Movie Predictions from Blogs** **Stanford University**
  Research Project *2008 - 09*
  This project dealt with trying to predict movie success using blog mentions. We extracted a comprehensive list of features that deal with movie references in blogs over the past year. We then identified the most useful features for predicting sales (overall, and in future weeks, given blog mentions until previous week) using correlation, clustering and time series techniques, and used these to achieve reasonable success at prediction.

- **Anonymity in Social Networks** **Stanford University**
  Research Project: Guide: Prof. Balaji Prabhakar *2008*
  We studied the problem of anonymizing social networks. We developed a new paradigm of anonymization (vertex randomization), and analyzed this theoretically on random graphs. We also analyzed how easy it is for an attacker to gain access to the entire social network graph, given various models of attack.

- **Robust 3-D Flow Complex and Visibility** **INRIA Sophia-Antipolis, France**
  Research Internship: Guide: Frederic Cazals and Sylvain Pion *Summer 2006*
  Contributed to the theory for the *first efficient and robust implementation* for the Three Dimensional Flow Complex. The Flow Complex is a geometric data structure with applications in surface reconstruction and shape segmentation. I also worked on the problem of visibility in a polygon formed out of circular arcs. We developed some new insights and a general Bentley-Ottman-like sweep algorithm for the same.

- **Termination of Linear Programs** **IIT Bombay, India**
  Senior Year Thesis: Guide: Prof Supratik Chakraborty *2006 - 07*
  I studied the problem of termination in my BTech thesis. We developed an algorithm to automatically determine if certain linear programs terminate, without actually executing the program.

- **Access Control Analysis** **Microsoft Research, Bangalore**
  Research Internship: Guide: Prasad Naldurg and Sriram Rajamani *Winter 2005*
  At Microsoft Research, I worked on the problem of using of formal methods to analyze access control permissions to evaluate the security of the Windows XP operating system. We located a number of security criteria based on information flows from authorized users and administrators to ordinary users and methods to check for their violation. Implemented a prototype tool for the same.

‘

Publications and Manuscripts

1. *Fuzzy Joins Using MapReduce*, Foto Afrati, Anish Das Sarma, David Menestrina, Aditya Parameswaran and Jeffrey D. Ullman, **28th International Conf. on Data Engineering (ICDE)**, Washington, Apr 2012
2. *Information Seeking: Convergence of Search, Recommendations and Advertising*, Georgia Koutrika, Hector Garcia-Molina and Aditya Parameswaran, **Communications of the ACM (CACM)**, Nov 2011
3. *Recommendation Systems with Complex Constraints: A Course Recommendation Perspective*, Aditya Parameswaran, Petros Venetis and Hector Garcia-Molina, **Transactions on Information Systems**, Volume 29(4), Nov 2011
4. *CrowdScreen: Algorithms for Filtering Data with Humans*, Aditya Parameswaran, Hector Garcia-Molina, Hyunjung Park, Neoklis Polyzotis, Aditya Ramesh and Jennifer Widom, Tech Report, Infolab, Sep 2011
5. *Optimal Schemes for Robust Web Extraction*, Aditya Parameswaran, Nilesh Dalvi, Hector Garcia-Molina and Rajeev Rastogi, **37th International Conf. on Very Large Data Bases (VLDB)**, Seattle, Sep 2011
6. *Human-assisted Graph Search: It's Okay to Ask Questions*, Aditya Parameswaran, Anish Das Sarma, Hector Garcia-Molina, Neoklis Polyzotis and Jennifer Widom, **37th International Conf. on Very Large Data Bases (VLDB)**, Seattle, USA, Sep 2011
7. *Answering Queries using Databases, Humans and Algorithms*, Aditya Parameswaran and Neoklis Polyzotis, **Conf. on Innovative Data Management (CIDR)**, Asilomar, USA, Jan 2011
8. *Evaluating, Combining and Generalizing Recommendations with Prerequisites*, Aditya Parameswaran, Hector Garcia-Molina and Jeffrey D. Ullman, **19th International Conf. on Information and Knowledge Management (CIKM)**, Toronto, Canada, Oct 2010
9. *Towards the Web of Concepts: Extracting Concepts from Large Datasets*, Aditya Parameswaran, Hector Garcia-Molina and Anand Rajaraman, **36th International Conf. on Very Large Data Bases (VLDB)**, Singapore, Sep 2010 (Invited to **Special Issue of VLDB Journal for VLDB 2010 Best Papers**)
10. *Recsplorer: Recommendation Algorithms Based on Precedence Mining*, Aditya Parameswaran, Georgia Koutrika, Benjamin Berkovitz and Hector Garcia-Molina, **ACM SIGMOD International Conf. on the Management of Data**, Indianapolis, USA, Jun 2010
11. *Synthesizing View Definitions from Data*, Anish Das Sarma, Aditya Parameswaran, Hector Garcia-Molina and Jennifer Widom, **13th International Conference on Database Theory (ICDT)**, Lausanne, Switzerland, Mar 2010
12. *Social Sites Research Through CourseRank*, Benjamin Berkovitz, Filip Kaliszan, Georgia Koutrika, Henry Liou, Aditya Parameswaran, Petros Venetis, Zahra Mohammadi Zadeh and Hector Garcia-Molina, Volume 38, **SIGMOD Record**, Dec 2009
13. *Recommendations with Prerequisites*, Aditya Parameswaran and Hector Garcia-Molina (Short Paper) **3rd ACM Conference on Recommender Systems (RecSys)**, New York, USA, Oct 2009
14. *Blogs as Predictors of Movie Success*, Eldar Sadikov, Aditya Parameswaran and Petros Venetis, **AAAI Conf. on Weblogs and Social Media (ICWSM)** May 2009, San Jose
15. *Anonymization of Social Networks*, Aditya Parameswaran and Ankur Taly, Tech Report, Stanford University, Aug 2008
16. *Robust Construction of the Three-dimensional Flow Complex*, Frederic Cazals, Aditya Parameswaran and Sylvain Pion, **ACM Symposium on Computational Geometry (SOCG)** 2008, Maryland, USA
17. *Termination of Linear Programs*, Aditya Parameswaran, BTech Thesis (Under the supervision of Supratik Chakraborty), June 2007, IIT Bombay
18. *A Vector approach to Ray Tracing for Reflection and Refraction*, Yashodhan Kanoria and Aditya Parameswaran, **Physics Education**, Jan-Mar 2006 issue, Vol 22, No.4

Talks

1. International Conf. on Very Large Data Bases (VLDB), Seattle, USA: September 11 (Two talks)
2. Crowd-Crowd Workshop, UC Berkeley: June 11
3. Conference on Innovative Database Research (CIDR), Asilomar, USA: January 11
4. Conference on Information and Knowledge Management (CIKM), Toronto, Canada: October 10
5. International Conf. on Very Large Data Bases (VLDB), Singapore: September 10
6. Microsoft Research DMX Group, Redmond, USA: August 10
7. SIGMOD International Conf. on Management of Data, Indianapolis, USA: June 10
8. International Conf. on Database Theory (ICDT), Lausanne, Switzerland: March 10
9. Yahoo Research Bangalore, India: September 09
10. UC Berkeley, Database Group Lunch Talk: March 09
11. Kosmix Corporation, Mountain View, USA: August 08

Academic Honors and Other Achievements

- IIT - Joint Entrance Examination (2003) - All India Rank of **34** out of nearly 150,000 students.
- National Science Olympiad (2003) - Ranked **4th in India** in this science examination.
- Institute Technical Citation - (2006-07) - Awarded the prestigious **Tech Citation** for excellence in technological extra-curricular activities over the four years at IIT - given to three students each year
- Indian National Physics Olympiad (2003) - Amongst the **top 25** gold medalists in the nation
- Technical Consultant (2007): Worked with the Scriptwriter and Director to **write dialogues for an Indian Film** ('Teen Patti' - released in Feb. 2010) with a budget of the order of tens of Millions of U.S. Dollars
- Member of the Institute Student Mentor body of IIT Bombay (2006): Mentored a group of 16 freshmen
- **Winner of several tech competitions** at IIT Bombay, including Science Journalism (twice), Tech Decathlon, Original Idea Presentation Contest, Yantriki (Robotics Competition)
- Passed Level one of the **Speech and Drama** examination conducted by Trinity College, London, with merit and the Levels 2 and 3 with distinction
- **Trained in singing** in the Carnatic Style (for two years) and in the Hindustani Style (for one year)

Programming Skills

- Languages: Python, C/C++, C#/Java, Scheme, SQL, Haskell, Prolog
- Scripting / Formatting: LaTeX, HTML, PHP, Awk, Sed, Bash
- Operating Systems: Linux (Ubuntu/Fedora), Windows 98/XP/Vista/7