

Aditya Parameswaran

Department of Computer Science
University of Illinois, Urbana-Champaign

E-mail: adityagp@illinois.edu
data-people.cs.illinois.edu

- Employment*
- ◇ **University of Illinois at Urbana-Champaign**, Urbana, IL.
Assistant Professor, Department of Computer Science. August 2014—
Affiliate Faculty, Institute for Genomic Biology (IGB). July 2016—
Affiliate Faculty, Beckman Interdisciplinary Institute for Advanced Science and Tech. July 2016—
 - ◇ **Massachusetts Institute of Technology**, Cambridge, MA.
Postdoctoral Researcher, Computer Science and Artificial Intelligence Lab. September 2013—August 2014.
 - ◇ **Microsoft Research New England**, Cambridge, MA.
Consulting Researcher. September 2013—January 2014.
- Education*
- ◇ **Stanford University**, Stanford, CA, USA
Ph.D. in Computer Science. September 2007—September 2013
Advisor: Hector Garcia-Molina; Title: *Human-Powered Data Management*
Winner: **ACM SIGMOD Jim Gray Dissertation Award**
Winner: **ACM SIGKDD Dissertation Award Runner Up**
Winner: **Stanford's Arthur Samuels Dissertation Award**
 - ◇ **IIT Bombay**, Mumbai, India
B.Tech in Computer Science. August 2003—August 2007
- Broad Interests*
- Visual and Interactive Data Analytics, Information Management, Crowdsourcing, Information Extraction
- Honors and Awards*
- ◇ **IEEE TCDE Early Career Award**, from the Technical Committee on Data Engineering, for my contributions to the field of databases in the first five years post PhD, Award citation: “*For developing new interactive tools and techniques that expand the reach of data analytics, enabling powerful data-driven discoveries by experts and non-experts alike.*”, 2017
 - ◇ **NSF CAREER Award**, 2017
 - ◇ Selections for **best papers** of the following conferences (awarded to ~5 papers per conference):
 - **AISTATS'17**, for my paper, titled “On the Interpretability of Conditional Probability Estimates in the Agnostic Setting”
 - **ICDE'16**, for my paper, titled “Interactive Data Exploration with Smart Drill-Down”
 - **ICDE'14**, for my paper, titled “Crowd-Powered Find Algorithms”
 - **KDD'12**, for my paper, titled “Active Sampling for Entity Matching”
 - **VLDB'10**, for my paper, titled “Towards the Web of Concepts: Extracting Concepts from Large Datasets”
 - ◇ Listed in “**Instructors rated as Excellent by students @ Illinois**” 2016
 - ◇ **Google Faculty Research Award** 2015
 - ◇ Awards for the dissertation “Human-powered Data Management”, including:
 - **ACM SIGMOD Jim Gray Dissertation Award**, 2013
 - **ACM SIGKDD Dissertation Award Runner Up**, 2013
 - **Stanford's Arthur Samuels Best Dissertation Award in Computer Science**, 2013
 - ◇ Selected as a **ACM Heidelberg Laureate** 2013
 - ◇ **Yahoo! Key Scientific Challenges Award, 2010**, for my proposal, “Information-Powered Recommendations”
- Relevant Expertise*
- My work centers on the design of *interactive or human-in-the-loop data analytics systems* by synthesizing techniques from multiple fields: **databases**, **data mining**, and **human computation**. I've published over **60 papers** in the top-tier venues of these fields with an **h-index of 25**. My research begins with a thorough exploration of the foundational principles, followed by the design of practical, scalable, and usable systems and algorithms.

- Research Themes
- ◊ **Visual Analytics [VA]:** Scalable visualization generation and recommendation on large datasets
 - ◊ **Interactive Analytics [IA]:** Interactively analyzing large volumes of data
 - ◊ **Optimized Human Computation [OC]:** Optimizing crowdsourcing for large-scale data processing
 - ◊ **Simplified Information Extraction [IE]:** Reducing human involvement in web information extraction
 - ◊ **Enhanced Recommendation Systems [RS]:** Using context to improve recommender systems
 - ◊ **Miscellaneous [MS]**

- Publications
- ◊ **Full Conference Papers** (Students I advise or co-advise are marked using a *.)
 1. [IA] S. Huang*, L. Xu*, J. Liu*, A. Elmore, **Aditya Parameswaran**. OrpheusDB: Bolt-on Versioning for Relational Databases. *VLDB '17: 43rd Int'l Conf on Very Large Data Bases*, Munich, Germany, 2017. Acceptance Rate: ~20%.
 2. [OC] A. Jain*, A. Das Sarma*, **Aditya Parameswaran**, and J. Widom. Understanding Workers, Developing Effective Tasks, and Enhancing Marketplace Dynamics: A Study of a Large Crowdsourcing Marketplace. *VLDB '17: 43rd Int'l Conf on Very Large Data Bases*, Munich, Germany, 2017. Acceptance Rate: ~20%.
 3. [VA] T. Siddiqui*, A. Kim*, J. Lee*, K. Karahalios, and **Aditya Parameswaran**. Effortless Visual Data Exploration with Zenvisage: An Expressive and Interactive Visual Analytics System. *VLDB '17: 43rd Int'l Conf on Very Large Data Bases*, Munich, Germany, 2017. Acceptance Rate: ~20%.
 4. [IE] T. Rekatsinas, M. Joglekar, H. Garcia-Molina, **Aditya Parameswaran**, and C. Re. SLIMFast: Guaranteed Results for Data Fusion and Source Reliability *SIGMOD '17: ACM SIGMOD Int'l Conf. on Management of Data*, Raleigh, USA, 2017. Acceptance Rate: ~20%.
 5. [MS] Y. Gao*, **Aditya Parameswaran**, and J. Peng. On the Interpretability of Conditional Probability Estimates in the Agnostic Setting (Oral Presentation). *AISTATS '17: Conf on Artificial Intelligence and Statistics*, Ft. Lauderdale, USA, 2017. Acceptance Rate: ~30%.
 6. [VA] T. Siddiqui*, J. Lee*, A. Kim*, E. Xue*, X. Yu*, S. Zou*, L. Guo*, C. Liu*, C. Wang*, K. Karahalios, and **Aditya Parameswaran**. Fast-forwarding to Desired Visualizations with Zenvisage. *CIDR '17: Conf. on Innovative Data Management (CIDR)*, Chaminade, USA, 2017.
 7. [IE] T. Siddiqui*, X. Ren, **Aditya Parameswaran**, and Jiawei Han. FacetGist: Collective Extraction of Document Facets in Large Technical Corpora. *CIKM '16: 25th Int'l Conf. on Information and Knowledge Management*, Indianapolis, USA, 2016. Acceptance Rate: 23%.
 8. [IA] M. Maddox, D. Goehring, A. Elmore, S. Madden, **Aditya Parameswaran**, and A. Deshpande. Decibel: The Relational Dataset Branching System. *VLDB '16: 42nd Int'l Conf on Very Large Data Bases*, New Delhi, India, 2016. Acceptance Rate: ~20%.
 9. [VA] M. Vartak*, S. Rahman*, S. Madden, **Aditya Parameswaran**, and N. Polyzotis. SeeDB: Efficient Data-Driven Recommendations to Support Visual Analytics. *VLDB '16: 42nd Int'l Conf on Very Large Data Bases*, New Delhi, India, 2016. Acceptance Rate: ~20%.
 10. [IA] Y. Gao* and **Aditya Parameswaran**. Squish: Near-optimal Compression for Archival of Relational Datasets. *KDD '16: 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, San Francisco, USA, 2016. Acceptance Rate: 6%.
 11. [OC] A. Das Sarma*, **Aditya Parameswaran**, and J. Widom. Towards Globally Optimal Crowdsourcing Quality Management. *SIGMOD '16: ACM SIGMOD Int'l Conf. on Management of Data*, San Francisco, USA, 2016. Acceptance Rate: 19%.
 12. [IA] M. Joglekar*, H. Garcia-Molina, and **Aditya Parameswaran**. Interactive Data Exploration with Smart Drill-down. *ICDE '16: 32nd Int'l Conf on Data Engineering*, Helsinki, Finland, 2016. Acceptance Rate: 25%. Invited to **Special Issue of TKDE for ICDE 2016 Best Papers**.
 13. [OC] A. Das Sarma*, A. Jain*, A. Nandi, **Aditya Parameswaran**, and J. Widom. Surpassing Humans and Computers with JellyBean: Crowd-Vision-Hybrid Counting Algorithms. *HCOMP '15: 3rd AAAI Int'l Conf. on Human Computation and Crowdsourcing*, San Diego, USA, 2015.
 14. [VA] M. Vartak*, S. Huang*, T. Siddiqui*, S. Madden, and **A. Parameswaran**. Towards Visualization Recommendation Systems, *DSIA '15: Workshop on Interactive Analysis*, Chicago, 2015.
 15. [IA] S. Bhattacharjee, A. Chavan, S. Huang*, A. Deshpande, and **Aditya Parameswaran**. Principles of Dataset Versioning: Exploring the Recreation/Storage Tradeoff. *VLDB '15: 41st Int'l Conf on Very Large Data Bases*, Hawaii, USA, 2015. Acceptance Rate: 21%.

16. [VA] A. Kim^{*}, E. Blais, **Aditya Parameswaran**, P. Indyk, S. Madden, and R. Rubinfeld. Rapid Sampling for Visualizations with Ordering Guarantees. *VLDB '15: 41st Int'l Conf on Very Large Data Bases*, Hawaii, USA, 2015. Acceptance Rate: 21%.
17. [OC] Y. Gao^{*} and **Aditya Parameswaran**. Finish Them!: Pricing Algorithms for Human Computation, *VLDB '15: 41st Int'l Conf on Very Large Data Bases*, Hawaii, USA, 2015. Acceptance Rate: 21%.
18. [OC] H. Zhuang^{*}, **Aditya Parameswaran**, D. Roth, and J. Han. Debiasing Crowdsourced Batches, *KDD '15: 21st ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, Sydney, Australia, 2015. Acceptance Rate: 20%.
19. [IA] A. Chavan, S. Huang^{*}, A. Deshpande, A. Elmore, S. Madden, and **Aditya Parameswaran**. Towards a Unified Query Language for Provenance and Versioning, *TAPP'15: 7th Int'l Conf. on Theory and Practice of Provenance*, Edinburgh, Scotland, 2015.
20. [IA] M. Joglekar^{*}, H. Garcia-Molina, **Aditya Parameswaran**, and C. Re. Exploiting Correlations for Expensive Predicate Evaluation, *SIGMOD '15: ACM SIGMOD Int'l Conf. on Management of Data*, Melbourne, Australia, 2015. Acceptance Rate: 26%.
21. [OC] M. Joglekar^{*}, Hector Garcia-Molina, and **Aditya Parameswaran**. Comprehensive and Reliable Crowd Assessment Algorithms, *ICDE '15: 31st Int'l Conf on Data Engineering*, Seoul, Korea, 2015. Acceptance Rate: 25%.
22. [IA] A. Bhardwaj, S. Bhattacharjee, A. Chavan, A. Deshpande, A.J. Elmore, S. Madden, and **Aditya Parameswaran**. DataHub: Collaborative Data Science & Dataset Version Management at Scale, *CIDR '15: Conf. on Innovative Data Management (CIDR)*, Asilomar, USA, 2015.
23. [OC] **Aditya Parameswaran**, S. Boyd, H. Garcia-Molina, A. Gupta, N. Polyzotis, and J. Widom. Optimal Crowd-Powered Rating and Filtering Algorithms, *VLDB '14: 40th Int'l Conf on Very Large Data Bases*, Hangzhou, China, 2014. Acceptance Rate: ~20%.
24. [OC] A. Das Sarma, **Aditya Parameswaran**, H. Garcia-Molina, and A. Halevy. Crowd-Powered Find Algorithms, *ICDE '14: 30th Int'l Conf on Data Engineering*, Chicago, USA, April 2014. Acceptance Rate: 20%. Invited to **Special Issue of TKDE for ICDE 2014 Best Papers**.
25. [OC] **Aditya Parameswaran**, M. H. Teh, H. Garcia-Molina, and J. Widom. DataSift: An Expressive and Accurate Crowd-Powered Search Toolkit, *HCOMP '13: 1st AAAI Int'l Conf. on Human Computation and Crowdsourcing*, Palm Springs, USA, 2013. Acceptance Rate: 30%.
26. [IE] **Aditya Parameswaran**, R. Kaushik, and A. Arasu. Efficient Parsing-based Search over Databases, *CIKM '13: 22nd Int'l Conf on Information and Knowledge Management*, Burlingame, USA, 2013. Acceptance Rate: 16.8%.
27. [OC] M. Joglekar, H. Garcia-Molina, and **Aditya Parameswaran**. Evaluating the Crowd with Confidence, *KDD '13: 19th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, Chicago, USA, 2013. Acceptance Rate: 17%.
28. [OC, IE] N. Dalvi, **Aditya Parameswaran**, and V. Rastogi. Minimizing Uncertainty in Pipelines, *NIPS '12: 25th Int'l Conf on Neural Information Processing Systems*, Tahoe, Nevada, USA, 2012.
29. [OC] **Aditya Parameswaran**, H. Park, H. Garcia-Molina, N. Polyzotis, and J. Widom. Deco: Declarative Crowdsourcing, *CIKM '12: 21st Int'l Conf on Information and Knowledge Management*, Maui, USA, 2012. Acceptance Rate: 13.4%.
30. [OC,IE] K. Bellare, S. Iyengar, **Aditya Parameswaran**, and V. Rastogi. Active Sampling for Entity Matching, *KDD '12: 18th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, Beijing, China, 2012. Acceptance Rate: 18%. Invited to **Special Issue of TKDD for KDD 2012 Best Papers**.
31. [OC] S. Guo, **Aditya Parameswaran**, and H. Garcia-Molina. So Who Won? Dynamic Max Discovery with the Crowd, *SIGMOD '12: ACM SIGMOD Int'l Conf. on the Management of Data*, Scottsdale, USA, 2012. Acceptance Rate: 17%.
32. [OC] **Aditya Parameswaran**, H. Garcia-Molina, H. Park, N. Polyzotis, A. Ramesh, and J. Widom. Crowd-Screen: Algorithms for Filtering Data with Humans *SIGMOD '12: ACM SIGMOD Int'l Conf. on the Management of Data*, Scottsdale, USA, 2012. Acceptance Rate: 17%.
33. [MS] F. Afrati, A. Das Sarma, D. Menestrina, **Aditya Parameswaran**, and J. D. Ullman. Fuzzy joins using MapReduce, *ICDE '12: 28th Int'l Conf. on Data Engineering*, Washington DC, USA, 2012. Acceptance Rate: 24%.

34. [IE] **Aditya Parameswaran**, N. Dalvi, H. Garcia-Molina, and R. Rastogi. Optimal Schemes for Robust Web Extraction, *VLDB '11: 37th Int'l Conf. on Very Large Data Bases*, Seattle, USA, 2011. Acceptance Rate: 18.1%.
35. [OC, IE] **Aditya Parameswaran**, A. Das Sarma, H. Garcia-Molina, N. Polyzotis, and J. Widom. Human-assisted Graph Search: It's Okay to Ask Questions, *VLDB '11: 37th Int'l Conf. on Very Large Data Bases*, Seattle, USA, 2011. Acceptance Rate: 18.1%.
36. [OC] **Aditya Parameswaran** and N. Polyzotis. Answering Queries using Databases, Humans, and Algorithms, *CIDR '11: Conf. on Innovative Data Management (CIDR)*, Asilomar, USA, 2011.
37. [RS] **Aditya Parameswaran**, H. Garcia-Molina, and J. D. Ullman. Evaluating, Combining, and Generalizing Recommendations with Prerequisites, *CIKM '10: 19th Int'l Conf. on Information and Knowledge Management*, Toronto, Canada, 2010. Acceptance Rate: 13%.
38. [IE] **Aditya Parameswaran**, H. Garcia-Molina, and A. Rajaraman. Towards the Web of Concepts: Extracting Concepts from Large Datasets, *VLDB '10: 36th Int'l Conf. on Very Large Data Bases*, Singapore, 2010. Acceptance Rate: 18.4%. Invited to **Special Issue of VLDB Journal for VLDB 2010 Best Papers**.
39. [RS] **Aditya Parameswaran**, G. Koutrika, B. Berkovitz, and H. Garcia-Molina. Recsplorer: Recommendation Algorithms Based on Precedence Mining, *SIGMOD '10: ACM SIGMOD Int'l Conf. on the Management of Data*, Indianapolis, USA, 2010. Acceptance Rate: 21%.
40. [IE] A. Das Sarma, **Aditya Parameswaran**, H. Garcia-Molina, and J. Widom. Synthesizing View Definitions from Data, *ICDT '10: 13th Int'l Conf. on Database Theory*, Lausanne, Switzerland, 2010. Acceptance Rate: 36%.
41. [RS] **Aditya Parameswaran** and H. Garcia-Molina. Recommendations with Prerequisites (Short Paper) *RecSys '09: 3rd ACM Conf. on Recommender Systems*, New York, USA, 2009. Acceptance Rate: 43%.
42. [MS] E. Sadikov, **Aditya Parameswaran**, and P. Venetis. Blogs as Predictors of Movie Success, *ICWSM '09: AAAI Conf. on Weblogs and Social Media*, San Jose, USA, 2009.
43. [MS] F. Cazals, **Aditya Parameswaran**, and S. Pion. Robust Construction of the Three-dimensional Flow Complex, *SOCG '08: ACM Symposium on Computational Geometry*, Maryland, USA, 2008. Acceptance Rate: 32%.

◊ **Journals**

44. [IA] M. Joglekar*, H. Garcia-Molina, and **Aditya Parameswaran**. Interactive Data Exploration with Smart Drill-down (Extended Version). *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, March 2017.
45. [VA] M. Vartak*, S. Huang*, T. Siddiqui*, S. Madden, and **A. Parameswaran**. Towards Visualization Recommendation Systems, *SIGMOD Record*, December 2016.
46. [OC] **Aditya Parameswaran**, A. Das Sarma*, and V. Venkataraman*. Optimizing Open-Ended Crowdsourcing: The Next Frontier in Crowdsourced Data Management. *IEEE Data Engineering Bulletin*, December 2016
47. [OC] H. Garcia-Molina, M. Joglekar*, A. Marcus, **Aditya Parameswaran**, and V. Verios. Challenges in Data Crowdsourcing. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, January 2016
48. [OC,IE] K. Bellare, S. Iyengar, **Aditya Parameswaran**, and V. Rastogi. Active Sampling for Entity Matching with Guarantees, *ACM Transactions on Knowledge Discovery from Data*, Volume 7(3), September 2013.
49. [OC] H. Park, R. Pang, **Aditya Parameswaran**, H. Garcia-Molina, N. Polyzotis, and J. Widom. An Overview of the Deco System: Data Model and Query Language; Query Processing and Optimization, *SIGMOD Record*, Volume 41, December 2012.
50. [RS] G. Koutrika, H. Garcia-Molina, and **Aditya Parameswaran**. Information Seeking: Convergence of Search, Recommendations, and Advertising, *Communications of the ACM (CACM)*, November 2011.
51. [RS] **Aditya Parameswaran**, P. Venetis, and H. Garcia-Molina. Recommendation Systems with Complex Constraints: A Course Recommendation Perspective, *ACM Transactions on Information Systems (TOIS)*, Volume 29(4), November 2011.
52. [RS] B. Berkovitz, F. Kaliszan, G. Koutrika, H. Liou, **Aditya Parameswaran**, P. Venetis, Z. Zadeh, and H. Garcia-Molina. Social Sites Research Through CourseRank, *SIGMOD Record*, Volume 38, December 2009.

◊ **Books**

53. [OC] A. Marcus and **Aditya Parameswaran**. Crowdsourced Data Management: Industry and Academic

Perspectives. *Foundations and Trends in Databases Series*, Vol. 6: No. 1-2, pp 1-161, Now Publishers, December 2015

◊ **Demo and Vision Papers**

54. [IA] L. Xu^{*}, S. Huang^{*}, S. Hui^{*}, A. Elmore, **Aditya Parameswaran**. OrpheusDB: A Light-weight Approach to Relational Dataset Versioning (Demo). *SIGMOD '17: ACM SIGMOD Int'l Conf. on Management of Data*, Raleigh, USA, 2017
55. [IA] M. Bendre^{*}, B. Sun^{*}, X. Zhou^{*}, D. Zhang^{*}, S. Lin^{*}, K. Chang, and **Aditya Parameswaran**. DataSpread: Unifying Databases and Spreadsheets (Demo) *VLDB '15: 41st Int'l Conf on Very Large Data Bases*, Hawaii, USA, 2015.
56. [IA] M. Joglekar^{*}, H. Garcia-Molina, and **Aditya Parameswaran**. Smart Drill-down: A New Data Exploration Operator (Demo). *VLDB '15: 41st Int'l Conf on Very Large Data Bases*, Hawaii, USA, 2015.
57. [IA] A. Bhardwaj, A. Deshpande, A. Elmore, D. Karger, S. Madden, **Aditya Parameswaran**, H. Subramanyam, E. Wu, and R. Zhang. Collaborative Data Analytics with Datahub (Demo). *VLDB '15: 41st Int'l Conf on Very Large Data Bases*, Hawaii, USA, 2015.
58. [MS] S. Koltani, S. Wang, **Aditya Parameswaran**. GeoHashViz: Interactive Analytics for Mapping Spatiotemporal Diffusion of Twitter Hashtags (Poster). *XSEDE '15*, USA, 2015.
59. [OC] A. Das Sarma^{*}, **Aditya Parameswaran**, and J. Widom. Optimal Worker Quality and Answer Estimates in Crowd-Powered Filtering and Rating (Short Paper). *HCOMP '14: 2nd Int' Conf. on Human Computation and Crowdsourcing*, Pittsburgh, USA, 2014.
60. [VA] **Aditya Parameswaran**, N. Polyzotis, and H. Garcia-Molina. SeeDB: Visualizing Database Queries Efficiently (Vision), *VLDB '14: 40th Int'l Conf on Very Large Data Bases*, Hangzhou, China, 2014.
61. [VA] M. Vartak^{*}, S. Madden, **Aditya Parameswaran**, and N. Polyzotis. SeeDB: Automatically Generating Query Visualizations (Demo), *VLDB '14: 40th Int'l Conf on Very Large Data Bases*, Hangzhou, China, 2014.
62. [OC] **Aditya Parameswaran**, M. H. Teh, H. Garcia-Molina, and J. Widom. DataSift: A Crowd-Powered Search Toolkit (Demo), *SIGMOD '14: ACM SIGMOD Int'l Conf. on the Management of Data*, Snowbird, USA, 2014.
63. [OC] H. Park, R. Pang, **Aditya Parameswaran**, H. Garcia-Molina, N. Polyzotis, and J. Widom. Deco: A System for Declarative Crowdsourcing (Demo), *VLDB '12: 38th Int'l Conf on Very Large Data Bases*, Istanbul, Turkey, 2012.
64. [OC] S. Guo, **Aditya Parameswaran**, and H. Garcia-Molina. So Who Won? Dynamic Max Discovery with the Crowd (Poster), *CrowdConf '11: 2nd Crowdsourcing Conference*, San Francisco, USA, 2011.

◊ **Preprints**

65. [IA] Y. Gao^{*}, S. Huang^{*}, **Aditya Parameswaran**. Navigating the Data Lake: Unsupervised Structure Extraction for Text-formatted Data Technical Report. December 2016
66. [IA] S. Rahman^{*}, M. Aliakbarpour, H. Kong^{*}, E. Blais, K. Karahalios, **Aditya Parameswaran**, and R. Rubinfeld. I've Seen Enough: Incrementally Improving Visualizations to Support Rapid Decision Making. Technical Report. November 2016
67. [OC] A. Jain^{*}, K. Goel^{*}, J. Young Seo^{*}, A. Kuznetsov^{*}, **Aditya Parameswaran**, H. Sundaram. It's a Matter of Perspective(s): Crowd-Powered Consensus Organization of Corpora. Technical Report. November 2016
68. [IA] M. Bendre^{*}, V. Venkataraman^{*}, X. Zhou^{*}, Kevin Chang, **Aditya Parameswaran**. Scaling up to Billions of Cells with DataSpread: Supporting Large Spreadsheets with Databases. Technical Report. November 2016
69. [IA] A. Kim^{*}, L. Xu^{*}, T. Siddiqui^{*}, S. Huang^{*}, S. Madden, **Aditya Parameswaran**. Speedy Browsing and Sampling with Needletail. Technical Report. November 2016
70. [OC] T. Rekatsinas, A. Deshpande, and **Aditya Parameswaran**. CrowdGather: Entity Extraction over Structured Domains. Technical Report. November 2014

71. [IA] L. Battle, T. Benson, **Aditya Parameswaran**, and E. Wu. Indexing Cost-Sensitive Prediction. Technical Report. August 2014
72. [IA] A. Kim*, S. Madden, and **Aditya Parameswaran**. NeedleTail: A System for Browsing Queries (Demo), Technical Report, April 2014
73. [OC] H. Park, **Aditya Parameswaran**, and J. Widom. Query Processing over Crowdsourced Data, Technical Report, August 2012.
74. [OC] A. Ramesh, **Aditya Parameswaran**, H. Garcia-Molina, and N. Polyzotis. Identifying Reliable Workers Swiftly, Technical Report, June 2012.

Major Software Releases

- ◇ **DataSpread**: A Spreadsheet-Database Hybrid. DataSpread has a spreadsheet frontend, and a database backend. DataSpread inherits the flexibility and ease-of-use of spreadsheets, as well as the scalability and power of databases, and scales to billions of cells seamlessly. 2016. URL: <http://dataspread.github.io>
- ◇ **OrpheusDB**: A relational dataset versioning system. OrpheusDB is built on top of standard relational databases, thus it inherits much of the same benefits of relational databases, while also compactly storing, tracking, and recreating versions on demand, all very efficiently. 2016. URL: <http://orpheus-db.github.io>
- ◇ **Zenvisage**: An ‘effortless’ data visualization tool. Zenvisage can automatically identify and recommend visualizations that match desired user patterns. The user can specify at a high level what they are looking for either via interactions or via a query language (ZQL), and the system will perform the necessary computation to identify these visualizations. 2016. URL: <http://zenvisage.github.io>
- ◇ **Squish**: A near-optimal structured data compression system. Squish identifies correlations between attributes to compress relational datasets both vertically as well as horizontally. 2016. URL: https://github.com/Preparation-Publication-BD2K/db_compress
- ◇ **Populace**: Software releases of optimized implementations of various crowdsourced data processing algorithms and systems. 2015. URL: <http://populace-org.github.io>

Funding

I have been part of teams that have brought in a total of over 12M in grant funding into Illinois; and over 3M as PI or Co-PI from a variety of federal and industry funding sources.

- ◇ Amazon Web Services **Research Credits**, 2017.
- ◇ Adobe Research **Faculty Research Gift**, 2016.
- ◇ Principal Investigator, **NSF CAREER Award** titled “Advancing Open-Ended Crowdsourcing: The Next Frontier in Crowdsourced Data Management”, 2017–2021.
- ◇ Principal Investigator, **NSF BIGDATA Award** titled “Bringing Interactive Data Management to Scientists, Analysts, and the Masses: A Holistic Unification of Spreadsheets and Databases”, 2016–2020.
- ◇ Principal Investigator, **Siebel Energy Institute** Seed Grant, “Data-driven Discovery of Resilient Energy Storage for Grid Applications”, Jointly awarded to Carnegie Mellon University and Illinois, 2016–17.
- ◇ Principal Investigator, **NIH BD2K** (Big Data to Knowledge) Supplementary Grant, “Piloting a Data Publication Service for the BD2K Commons”, Jointly awarded to U Chicago and Illinois, 2015–16.
- ◇ Principal Investigator, **NSF IIS Medium Award** titled “DataHub: Collaborative Dataset Management for Data Science”, Jointly awarded to Illinois, Maryland, and MIT, 2015–2018.
- ◇ Google **Faculty Research Award**, 2015.
- ◇ Investigator, **NIH BD2K** (Big Data to Knowledge) Centers of Excellence titled “KnowEng, a Scalable Knowledge Engine for Large Scale Genomic Data”, Jointly awarded to Illinois and Mayo Clinic, 2014–2018.
- ◇ Google Cloud Credit Award, 2014.

Industry Internships

- ◇ **Yahoo! Research**, Web Information Management Group, Santa Clara, CA Summer 2011
- ◇ **Microsoft Research**, Data Management, Mining and Exploration Group, Redmond, WA Summer 2010
- ◇ **Yahoo! Research**, Information Extraction Group, Bangalore, India Winter 2009
- ◇ **Kosmix**, Mountain View, CA Summer 2008

- ◇ **INRIA**, Geometric Algorithms Group, Sophia-Antipolis, France Summer 2006
 - ◇ **Microsoft Research**, Formal Methods Group, Bangalore, India Winter 2005
- Mentorship*
- ◇ PhD Students (Current Advising or Co-advising)
 - Illinois: Doris Xin, 2016– (Winner: **NSF GRFP**);
 - Illinois: Liqi Xu, 2015–;
 - Illinois: Stephen Macke, 2015– (Winner: **NSF GRFP**);
 - Illinois: Tana Wattanawaroon, 2015– (Winner: **Outstanding TA Award**);
 - Illinois: Mangesh Bendre, 2015– (Co-advisor: Kevin Chang);
 - Illinois: Sajjadur Rahman, 2014–;
 - Illinois: Silu Huang, 2014– (Winner: **3M Fellowship**, Winner: **MSR Fellowship**—the first Illinois student winner in 7 years.);
 - Illinois: Yihan Gao, 2013– (Winner: **Richard T. Cheng Fellowship**);
 - Stanford: Akash Das Sarma, 2012– (Co-advisor: Jennifer Widom);
 - ◇ PhD Students (Informal or Past Advising)
 - MIT: Albert Kim, 2013–16;
 - Illinois: Himel Dev, 2015–16;
 - MIT: Manasi Vartak, 2013–15;
 - Stanford: Manas Joglekar, 2012–16 (First employment: Google);
 - ◇ MS Students (Advising or Co-advising)
 - Illinois: Shreya Rajpal, 2016–; Vipul Venkataraman, 2015– (Winner: **Siebel Scholarship, Outstanding TA Award**); Tarique Siddiqui, 2014–16 (Winner: **Siebel Scholarship**, First employment: PhD student at Illinois); Sili Hui, 2014–16 (First employment: Medallia); Edward Xue, 2014–; Chaoran Wang, 2015–; Ayush Jain 2014–16 (Winner: **Muroga Fellowship**, First employment: Google);
 - Stanford: Ming Han Teh, 2012–14 (Winner: **Stanford’s MS Thesis Award**, First employment: Microsoft); Aditya Ramesh, 2011–12 (First employment: LinkedIn);
 - ◇ Undergraduate Students
 - Illinois: Bangqi Wang, 2016–; Soumitri Vadali, 2016–; Paul Luh, 2016–; Kejia Jiang, 2015–; Kelly Mack, 2015–; Abhishek Nigam, 2016– (First employment: Jump Trading); Andrew Kuznetsov, 2014– (Winner: **Illinois SURF Research Award**, 3rd place at the **Engineering Open House**); Xiaofu Yu, 2015–16; Lijin Guo, 2015–16; Gary Luo, 2015–16 (First employment: Yahoo! Champaign); Joon Seo, 2014– (First employment: Google); Yulun Du, 2014–15 (First employment: Graduate student at Carnegie Mellon); Ding Zhang, 2014–16 (First employment: Graduate student at Carnegie Mellon); Xinyan (Paul) Zhou, 2014–16 (First employment: Graduate student at Illinois); Anurag Choudhury, 2015–16 (First employment: Microsoft)
- External Community Activities*
- ◇ **Tutorials:**
 - Invited 3-hour Tutorial at the HCOMP (Human Computation and Crowdsourcing) Conference 2016, titled “*Crowdsourced Data Management: Industry and Academic Perspectives*”, with Adam Marcus, 2016.
 - ◇ **Working Groups:**
 - NIH BD2K Commons Working Group, 2015–
 - NIH BD2K Machine Learning Working Group, 2016–
 - ◇ **Editorships:**
 - Associate Editor, SIGMOD Record, December 2014–
 - ACM XRDS Magazine (Issue on “Big Data”), with 15 invited articles from experts, 2012
 - ◇ **Chairing:**
 - Chair, Human-in-the-loop Data Analytics Workshop (HILDA) at SIGMOD, 2017.
 - Chair, SIGMOD Undergraduate Research Competition, 2016.
 - Area Chair, SIGMOD 2017

- Chair, Workshops and Tutorials, 1st HCOMP (Human Computation Conference) 2013
- ◇ **Award Committees:**
 - SIGMOD Jim Gray Dissertation Award Committee, 2015
 - SIGMOD Best Demo Award Committee, 2014
 - PURE Undergraduate Research Award Committee, 2014
- ◇ **Panel Participation:**
 - Plenary Panel “Will AI Eat Us All?” at VLDB, 2016
 - Enterprise Intelligence Workshop Panel at KDD, 2016
- ◇ **Review Panels:**
 - NSF IIS-III Databases Panel, April 2017
 - NSF IIS-III Databases Panel, April 2016
 - NSF IIS-III Data Mining Panel, June 2015
- ◇ **Program Committee Member:**
 - ICDE Demo 2018
 - SIGMOD Demo 2017
 - HCOMP 2017, 2014, 2012
 - VLDB 2016, 2015, 2014
 - KDD 2015
 - SIGMOD 2015, 2014
 - SOCC 2014
 - WSDM 2014
 - ICDE 2014
 - EDBT 2014
 - WWW 2014
- ◇ **Journal Reviewer:** VLDB J., 2015; J. of Data and Information Quality, 2014; SIGMOD Record, 2014; ACM TODS, 2014; Information Systems, 2014, 2013; CACM, 2013; SIAM J. of Computing, 2013; IEEE TKDE, 2016-15, 2010-13; ACM TOIS, 2012; J. of Web Semantics 2011
- ◇ **Workshop Participation:**
 - BIRS-CMO Workshop on “Theory and Models of Crowds and Networks”, Oaxaca, Mexico, 2016

Invited Talks

- ◇ IBM TJ Watson Invited Lecture: December 2016
- ◇ Northwestern EECS Distinguished Lecture: November 2016
- ◇ Midwest BigData Workshop: October 2016
- ◇ U. Illinois BigData Day: October 2016
- ◇ BIRS-CMO Workshop on Theory of Crowds and Networks: August 2016
- ◇ SIGKDD EI (Enterprise Intelligence) Workshop Keynote: August 2016
- ◇ National Center for Supercomputing Applications: August 2016
- ◇ University of Texas-Austin: November 2015
- ◇ University of Michigan: October 2015
- ◇ NIH Center Meeting: August 2015
- ◇ Google Mountain View Invited Talk: September 2014
- ◇ SIGKDD Dissertation Award Talk: August 2014
- ◇ SIGKDD IDEA (Interactive Data Exploration and Analysis) Workshop Invited Keynote: August 2014
- ◇ SIGMOD Jim Gray Award Keynote: June 2014
- ◇ UMass Amherst Invited Talk: April 2014
- ◇ MIT Data Analytics Workshop: April 2014
- ◇ INFORMS Conference Invited Talk: October 2013
- ◇ LinkedIn Data Science: August 2013

- ◇ UMichigan CS Colloquium: April 2013
- ◇ USC CS Colloquium: April 2013
- ◇ CMU SCS Colloquium: March 2013
- ◇ Illinois CS Colloquium: March 2013
- ◇ UC San Diego CS Colloquium: March 2013
- ◇ Princeton CS Colloquium: March 2013
- ◇ Harvard CS Colloquium: March 2013
- ◇ Georgia Tech CS Colloquium: March 2013
- ◇ University of Maryland CS Colloquium: February 2013
- ◇ University of Chicago CS Colloquium: February 2013
- ◇ Columbia University CS Colloquium: February 2013
- ◇ UC Santa Cruz Database Group Invited Talk: February 2013
- ◇ Duke University Invited Talk: January 2013
- ◇ Crowd-Crowd Workshop, Stanford: January 2013
- ◇ Amazon Machine Learning Group Invited Talk, Bangalore, India: September 2012
- ◇ IBM Research Invited Talk (HCI Group), Almaden, USA: July 2012
- ◇ MIT EECS Colloquium: April 2012
- ◇ Univ. of Washington CS Colloquium: February 2012
- ◇ IBM Research Invited Talk, Almaden, USA: March 2011
- ◇ Crowd-Crowd Workshop, UC Berkeley: June 2011
- ◇ Microsoft Research DMX Group, Redmond, USA: August 2010
- ◇ Yahoo! Research Invited Talk, Bangalore, India: September 2009
- ◇ UC Berkeley, Database Group Lunch Talk: March 2009
- ◇ Kosmix Corporation, Mountain View, USA: August 2008