Saswat Padhi

Curriculum Vitae • Last modified on February 18, 2024

spadhi@google.com • Google SJC-TM-2 • San Jose, CA 95126
saswatpadhi@protonmail.com • https://saswat.padhi.me/

Research Interests

I am a backend generlist, primarily interested in software infrastructure. My research spans program analysis, software verification, program synthesis, and machine learning. More recently, my work has been focused on performance and reliability analysis of large-scale industrial software systems.

Education

University of California, Los Angeles Sep 2014 – Jun 2020
(UCLA) *Los Angeles, CA*

- Philosophiae Doctor (Ph.D.), Computer Science · CGPA: 3.8/4.0
- *Dissertation:* Data-Driven Learning of Invariants and Specifications · *Advisor:* Prof. Todd Millstein

Indian Institute of Technology, Bombay Aug 2010 – Apr 2014
(IIT-B) *Mumbai, India*

- Bachelor of Technology (B.Tech.), Computer Science and Engineering with Honors · CPI: 8.9/10.0
- *Undergraduate Thesis:* Lazy Static Slicing of Functional Programs · *Advisor:* Prof. Amitabha Sanyal

D.A.V. Public School, Pokhariput 2004 – 2009
(DAV-Pkt) *Bhubaneswar, India*

- 2009 All India Senior School Certification Examination in Science · 96.0 %
- 2007 All India Secondary School Examination · 98.0 %

Employment

Google LLC Sep 2022 – Present
Senior Software Engineer *San Jose, CA*

- ChromeOS Platform Engineering, reporting to Sunil K. Bhasin

Amazon Inc. Aug 2020 – Sep 2022
Applied Scientist II *Boston, MA*

- Automated Reasoning Group (ARG), reporting to Dr. Michael Whalen and Dr. Jim Grundy

Microsoft Corp. Oct 2017 – Aug 2018
(Part-Time) Research Software Development Engineer *Remote (US)*

- Research in Software Engineering (RiSE) group, reporting to Dr. Ben Zorn

Visiting Positions

Princeton University Apr 2019 – Jun 2019
Visiting Research Collaborator *Princeton, NJ*

- Collaboration with Prof. David Walker in the Programming Languages Group
- Resulting paper, *Data-Driven Inference of Representation Invariants*, at PLDI 2020

Microsoft Research Lab - India Sep 2018 – Mar 2019
Ph.D. Research Intern *Bengaluru, India*

- Systems Research group, collaborating with Dr. Rahul Sharma
- Resulting paper, *Overfitting in Synthesis: Theory and Practice*, at CAV 2019

Microsoft Research Jun 2017 – Jun 2018
Ph.D. Research Intern *Redmond, WA*

- Research in Software Engineering (RiSE) group, with Dr. Ben Zorn and Dr. Rishabh Singh
- Developed a neural (CNN-based) approach for identifying tables (data frames) in Excel sheets

Microsoft Jun 2016 – Dec 2016
Ph.D. Research Intern *Redmond, WA*

- Program Synthesis using Examples (PROSE) group, with Dr. Sumit Gulwani
- Resulting paper, *FlashProfile: A Framework for Synthesizing Data Profiles*, at OOPSLA 2018

Google Inc. May 2013 – Jul 2013
Summer Intern *Mountain View, CA*

- Technical Infrastructure (TI) group, reporting to Smeeta Jalan
- Worked on test automation for Google's Borg and Omega cluster-management systems

Technische Universität Braunschweig (TU-Br) May 2012 – Jul 2012
Summer Research Intern *Braunschweig, Germany*

- Institut für Informationssysteme (IFIS), advised by Prof. Dr. Wolf-Tilo Balke
- Worked on a comparative analysis of bibliometric vs semantic measures for estimating topical similarity of scientific publications (primarily the PubMed corpus)

Honors and Awards

2020	*ACM SIGPLAN Distinguished Paper* award at the 41st PLDI Conference for "*Data-Driven Inference of Representation Invariants*" paper
2020	*Outstanding Research in Computer Science* award from UCLA
2019 – 2020	*Dissertation-Year Fellowship* from UCLA
2018	Gold medal at the *Federated Logic Conference (FLoC) Olympic Games* in London, UK for multiple consecutive victories in the annual SyGuS Competition
2017, 2018	Winner of the annual *Syntax-Guided Synthesis (SyGuS) Competition*
2017 – 2019	*Ph.D. Fellowship* from Microsoft Research
2013	Winner of the *Prezi Scale Contest* hosted by HackerRank
2010 – 2014	FIITJEE Scholarship for achieving *All-India Rank 43 in IIT-JEE*
2008	*Kishore Vaigyanik Protsahan Yojana (KVPY) Fellowship* from the Department of Science and Technology, Government of India
2008, 2007	Selected (top 25 across India) in the Indian National Astronomy Olympiad (INAO)
2007	*National Talent Search Examination (NTSE) Scholarship* from the National Council of Education Research and Training, Government of India
2007	Double gold medal from the University of New South Wales (UNSW), Australia for highest marks in Maths and CS in the International Assessment for Indian Schools (IAIS)
2005	Silver medal (All-India Rank 2) in the National Science Olympiad (NSO), India

Patent Grants and Applications

Syntactic profiling of alphanumeric strings.
Sumit Gulwani, Prateek Jain, Daniel Perelman, Saswat Padhi, Oleksandr Polozov

- 2021 Grant: US 11210327 B2 · 2019 Grant: US 10394874 B2
- An algorithm for generating a *syntactic profile* for a collection of strings: a set of *RegExp* like patterns, each of which describes the structure of a subset of strings.
- Implemented in the `Matching.Text` library, publicly available as part of Microsoft PROSE SDK.

Record profiling for dataset sampling.
Daniel Simmons, Kevin Grealish, Sumit Gulwani, Ranvijay Kumar, Kevin Ellis, Saswat Padhi

- 2020 Grant: US 10846298 B2
- An algorithm for sampling large datasets based on their syntactic profiles.
- Implemented in the `Matching.Text` library, publicly available as part of Microsoft PROSE SDK.

Systems, methods, and computer-readable media for improved table identification using a neural network.
Ben Zorn, Marc Brockschmidt, Pallavi Choudhury, Oleksandr Polozov, Rishabh Singh, Saswat Padhi

- 2020 Application: US 20200019603 A1
- A neural (CNN-based) algorithm for identifying tables (data frames) within Excel spreadsheets.
- Initially implemented as a Microsoft Excel add-on.

Publications

Journal Articles

1. *FlashProfile: A Framework for Synthesizing Data Profiles.* Saswat Padhi, Prateek Jain, Daniel Perelman, Oleksandr Polozov, Sumit Gulwani, and Todd Millstein. In the proceedings of the ACM on Programming Languages (**PACMPL**) Vol. 2, Object-Oriented Programming, Systems and Applications (**OOPSLA - 2018**) issue, Article 150, 2018. 28 pages. DOI: 10.1145/3276520.

Conference Proceedings

1. *Data-Driven Inference of Representation Invariants.* Anders Miltner, Saswat Padhi, Todd Millstein, and David Walker. In the proceedings of the 41st ACM SIGPLAN Conference on Programming Language Design and Implementation (**PLDI - 2020**), London, UK, June 15–20, 2020. 19 pages. DOI: 10.1145/3395638.
   (ACM SIGPLAN Distinguished Paper)
2. *Overfitting in Synthesis: Theory and Practice.* Saswat Padhi, Todd Millstein, Aditya Nori, and Rahul Sharma. In the proceedings of the 31st International Conference on Computer Aided Verification (**CAV - 2019**), Part I, New York City, NY, USA, July 15-18, 2019. 19 pages. Springer International Publishing. DOI: 10.1007/978-3-030-25540-4_17
3. *A Static Slicing Method for Functional Programs and Its Incremental Version.* Prasanna Kumar K., Amitabha Sanyal, Amey Karkare, and Saswat Padhi. In the proceedings of the 28th International Conference on Compiler Construction (**CC - 2019**), Washington, DC, USA, February 16–17, 2019. 12 pages. DOI: 10.1145/3302516.3307345
4. *Data-Driven Precondition Inference with Learned Features.* Saswat Padhi, Rahul Sharma, and Todd Millstein. In the proceedings of the 37th ACM SIGPLAN Conference on Programming Language Design and Implementation (**PLDI - 2016**), Santa Barbara, CA, USA, June 13–17, 2016. 19 pages. DOI: 10.1145/2908080.2908099

Workshop Papers

1. *Predicting User Experience on Laptops from Hardware Specifications.* Saswat Padhi, Sunil Bhasin, Udaya Kiran Ammu, Alex Bergman, and Allan Knies. In the proceedings of the NeurIPS 2023 Workshop on Machine Learning for Systems (**ML4Sys @ NeurIPS - 2023**), New Orleans, LA, USA, December 16, 2023. 9 pages.
   (Invited for Oral Spotlight Presentation)
2. *OASIS: ILP-Guided Synthesis of Loop Invariants.* Sahil Bhatia, Saswat Padhi, Nagarajan Natarajan, Rahul Sharma, and Prateek Jain. In the proceedings of the NeurIPS 2020 Workshop on Computer-Assisted Programming (**CAP @ NeurIPS - 2020**), December 12, 2020. 5 pages.

Industrial Experience Papers

1. *Automated Analyses of IoT Event Monitoring Systems.* Andrew Apicellii, Sam Bayless, Ankush Das, Andrew Gacek, Dhiva Jaganathan, Saswat Padhi, Vaibhav Sharma, Michael W. Whalen, and Raveesh Yadav. In the proceedings of the 35th International Conference on Computer Aided Verification (**CAV - 2023**), Part I, Paris, France, July 17-22, 2023. 13 pages. Springer International Publishing. DOI: 10.1007/978-3-031-37706-8_2

Theses

1. *Data-Driven Learning of Invariants and Specifications.* Saswat Padhi. Doctoral Dissertation, 2020. University of California, Los Angeles, CA. Supervised by Prof. Todd Millstein. ProQuest ID: Padhi_ucla_0031D_19057. DOI: https://escholarship.org/uc/item/3k89r896

Technical Reports

1. *The SyGuS Language Standard Version 2.1.* Saswat Padhi, Elizabeth Polgreen, Mukund Raghothaman, Andrew Reynolds, and Abhishek Udupa. Presented at SYNT 2021.

Invited Presentations

Data-Driven Precondition Inference

- Sep 2020: Amazon Web Services, Inc.
- Jan 2017: Microsoft Research Lab, Bengaluru, India.
- Jun 2016: 37th ACM SIGPLAN PLDI Conference, Santa Barbara, CA, USA.
- May 2016: University of California at Berkeley, CA, USA.
- May 2016: Software Lunch Talk, Stanford University, CA, USA.
- Dec 2015: 15th SoCal PLS Workshop, Pomona College, CA, USA.

Synthesizing Pattern-Based Data Profiles

- Nov 2018: 18th ACM SIGPLAN SPLASH-OOPSLA Conference, Boston, MA, USA.
- Oct 2018: Microsoft Research Lab, Bengaluru, India.
- Dec 2016: Microsoft Research, Redmond, WA, USA.

Overfitting in Program Synthesis

- Jul 2019: 31st CAV Conference, New York City, NY, USA.

ML-Based User Experience Prediction

- Dec 2023: 37th NeurIPS Conference, New Orleans, LA, USA.

Professional Service

Organizing Committee Member

- 2019 – 2021 Syntax-Guided Synthesis Competition (SyGuS-Comp)

Program / Review Committee Member

- 2024 HCVS Workshop at the European Joint Conferences on Theory & Practice of Software
- 2022 HCVS Workshop at the European Joint Conferences on Theory & Practice of Software
- 2021 ACM SIGPLAN Conference on Programming Languages Design and Implementation (PLDI)
- 2020 ACM SIGPLAN Conference on Programming Languages Design and Implementation (PLDI)
- 2021 SYNT Workshop at the International Conference on Computer-Aided Verification (CAV)
- 2019 DebugML Workshop at the International Conference on Learning Representations (ICLR)

Artifact Evaluation Committee Member

- 2020 ACM SIGPLAN Conference on Principles of Programming Languages (POPL)
- 2019 Static Analysis Symposium (SAS)
- 2019 ACM Conference on Object-Oriented Programming, Systems, and Applications (OOPSLA)
- 2018 ACM Conference on Object-Oriented Programming, Systems, and Applications (OOPSLA)

Invited External Reviewer

- 2022 International Conference on Foundations of Software Science and Computation Structures
- 2021 IEEE Transactions on Software Engineering (TSE)
- 2019 International Conference on Computer-Aided Verification (CAV)
- 2019 ACM SIGSOFT India's Innovations in Software Engineering Conference (ISEC)

Mentoring Experience

- Dr. Akshay Utture, intern at AWS during doctoral study at UCLA. Now at Uber Inc.
  (Summer 2022) Proving memory-safety of the task queue implementation in the FreeRTOS kernel.
- Dr. Imtiaz Kareem, intern at AWS during doctoral study at Purdue University.
  (Summer 2021, co-mentored with Dr. Vaibhav Sharma) Type checking for AWS-internal DSLs.
- Long Pham, intern at AWS during doctoral study at CMU.
  (Summer 2021) Adding termination proofs to the C Bounded Model Checker (CBMC).
- Dr. Aalok Thakkar, intern at AWS during doctoral study at UPenn. Now at Aptos Labs.
  (Summer 2021) Adding history variables to the C Bounded Model Checker (CBMC).
- Pamina Georgiou, intern at AWS during doctoral study at TU-Wein.
  (Fall 2020) Proving correctness of AWS projects in C using the C Bounded Model Checker (CBMC).
- Adam Stein, undergrad at UCLA. Now Ph.D. candidate at UPenn.
  (2019 – 2020) Adding Bit Vector theory to the LoopInvGen synthesizer.
- Sahil Bhatia, Research Fellow at Microsoft. Now Ph.D. candidate at UC Berkeley.
  (2018 – 2019) Synthesizing loop invariants using integer linear programming (ILP).
- Brett Chalabian, M.S. student at UCLA. Now at Google.
  (2017 – 2019) Adding user-specified ad-hoc features in the LoopInvGen synthesizer.
- Zhouheng (Jeffrey) Sun, undergrad at UCLA. Now at Apple.
  (2017 – 2019) Adding higher-order features to the LoopInvGen synthesizer.

Teaching Experience

Programming Languages CS 131
Graduate Teaching Assistant *University of California, Los Angeles*

- Spring 2016 · Instructor: Prof. Todd Millstein
- Fall 2014 · Instructor: Prof. Todd Millstein

Abstractions and Paradigms in Programming CS 152
Undergraduate Teaching Assistant *Indian Institute of Technology, Bombay*

- Spring 2014 · Instructor: Prof. Amitabha Sanyal

Computer Programming and Utilization Fall 2011 · Fall 2013
Undergraduate Teaching Assistant *Indian Institute of Technology, Bombay*

- Fall 2013 · Instructor: Padma Shri Prof. Deepak B. Phatak
- Fall 2011 · Instructor: Padma Shri Prof. Deepak B. Phatak

Software Releases

pseudocode.js 2020 – Present
[https://github.com/SaswatPadhi/pseudocode.js] *MIT License*

- Current lead developer; originally authored by Tate Tian.
- A JavaScript library that accurately typesets LaTeX-style pseudocodes to HTML.

LoopInvGen 2016 – 2020
[https://github.com/SaswatPadhi/LoopInvGen] *MIT License*

- Lead developer; co-developed with Todd Millstein and Rahul Sharma.
- Infers necessary preconditions and sufficient loop invariants, fully automatically.
- Won SyGuS-Comp 2017, 2018. Second place in SyGuS-Comp 2019.

Personal Information

- Born in March, 1992.
- Citizen of the Republic of India.
- Married and one child.
- Languages spoken: Odiya (native), English (fluent), Hindi (fluent), Sanskrit (basic).