

Saswat PADHI

Computer Science (PhD), UCLA

3440 Boelter Hall
University of California Los Angeles, CA 90024
padhi@cs.ucla.edu

<http://saswatpadhi.github.io/>
[saswatpadhi](#) · [saswatpadhi](#) · [saswat.padhi](#)

Interests

Data-Driven Learning, Program Synthesis, Program Verification

Education

- 2014 – Present **MS / PhD in Computer Science**,
University of California - Los Angeles, CGPA: 3.81 / 4.0 (advisor: Prof. Todd Millstein).
- 2010 – 2014 **B. Tech. in Computer Science and Engineering (Honors)**,
Indian Institute of Technology - Bombay, CGPA: 8.85 / 10.0.
- 2009 **All India Senior School Certification Examination in Science**,
D.A.V. Public School - Pokhariput, Bhubaneswar, Score: 96.00% (State Rank 4th).
- 2007 **All India Secondary School Examination (Matriculation)**,
D.A.V. Public School - Pokhariput, Bhubaneswar, Score: 98.00% (State Rank 3rd).

Publications

- PLDI 2016 Data-Driven Precondition Inference with Learned Features**
ACM SIGPLAN Conference on Programming Language Design and Implementation
[Saswat Padhi](#) · [Rahul Sharma](#) · [Todd Millstein](#)
Artifact <https://github.com/SaswatPadhi/PIE>

Awards and Honors

- 2017 – 2019 Microsoft Research PhD Fellowship
- 2013 Winner of Prezi Scale Contest
- 2010 – 2014 FIITJEE Scholarship for national rank 43 in IIT-JEE examination
- 2008 KVPY Fellowship (Dept. of Science and Technology, India)
- 2007 National Talent Search Scholarship (NCERT, Govt. of India)
- 2005 Silver medallist (national rank 2) in National Science Olympiad

Research Experience

- Microsoft (Intern) Project Jura: Generating Insights from Spreadsheet Data**
with *Dr. Ben Zorn*, RiSE Group, Microsoft Research | Redmond, WA Summer '17
I built a framework for extracting *tabular data* from within spreadsheets, and generating a variety of insights on them. I used a CNN (convolutional neural network) to predict likely table corners, and reconstruct tables from them. Using clever data augmentation techniques, I was able to train the CNN with extremely few labelled data. Within a table, we identify columns that are likely to be derived from other columns, and suggest likely Excel operators to compute them. I am currently working on synthesizing full formulas for deriving a column from other columns within a table.

Microsoft **FlashProfile: Synthesizing Syntactic (Pattern) Data Profiles**
(Intern) with *Dr. Sumit Gulwani*, PROSE, Data Group | Redmond, WA *Summer '16 – Fall '16*

I designed a technique for learning a *profile* for a string dataset, as a set of syntactic patterns that describe it. It builds a hierarchical clustering of the dataset using the complexity of the pattern required to describe a set of strings as a measure of their dissimilarity, and allows users to interactively “cut” the hierarchy tree at a desired granularity. Finally, clusters are labelled with patterns that describe the contained strings. I used the PROSE framework for efficiently learning patterns describing a set of strings.

UCLA **PIE: Feature Learning for Data-Driven Specification Inference**
(Graduate RA) with *Prof. Todd Millstein*, PL and SE Lab *Fall '14 – Winter '16*

We address a key challenge faced by prior data-driven specification inference techniques – defining a good search space for the logical formula to be learned. By combining traditional program analysis techniques with insights from machine learning, we present a novel technique to infer rich preconditions for programs, given a desired postcondition. We have also shown that our feature learning technique naturally leads to a new, more expressive and less onerous form of automatic full program verification.

IIT Bombay **Lazy Static Slicing of Functional Programs**
(Undergraduate RA) with *Prof. Amitabha Sanyal*, Compilers Group *Fall '13 – Spring '14*

I was investigating static analysis techniques for backward slicing functional programs – extracting the expressions involved in the computation of the final result. We proposed a lazy technique for computing slices of first order functional programs, which recursively transformed a slicing *demand* on an expression to a demand on its sub-expressions.

TU Braunschweig **Bibliometric vs Semantic Measures for Estimating Topical Similarity**
(Intern) with *Prof. Wolf-Tilo Balke*, Institut für Informationssysteme *Summer '12*

Using knowledge bases and ideas from information systems, I improved searches on semi-structured documents. The project aimed at a comparative analysis of bibliometric vs semantic similarity measures in estimating topical similarity of scientific literature.

Professional Experience

Google **Automating testing & maintenance of Omega components**
(Intern) with *Smeeta Jalan*, Technical Infrastructure Team | Mountain View, CA *Summer '13*

I was working with the Borg-Omega teams (Google’s cluster management systems) on automating various aspects of testing their systems. The goal was to automatically update and benchmark Omega components with respect to Borg ones, using real traffic data.

Teaching Experience

UCLA **Programming Languages** *Fall '14 • Spring '16*
Graduate Teaching Assistant

IIT Bombay **Abstractions and Paradigms in Programming** *Spring '14*
Undergraduate Teaching Assistant

IIT Bombay **Computer Programming and Utilization** *Fall '13 • Fall '11*
Undergraduate Teaching Assistant