

Vu Le

One Microsoft Way
Redmond, WA 98052
■ +1 530 304 8376
✉ levu@microsoft.com
🌐 <http://vuminhle.com>

Research Interests

Programming languages and software engineering: program synthesis, static and dynamic program analysis, and testing. In particular, I develop novel and practical techniques to help improve the quality of critical software (e.g., compilers and database engines) and make programming accessible for end users.

Education

- 2009-2015 **Ph.D., Computer Science**, University of California, Davis, USA.
Thesis: *Program Synthesis for Empowering End Users and Stress-Testing Compilers*.
Advisors: Prof. Zhendong Su and Dr. Sumit Gulwani.
- 2001-2006 **BEng, Computer Science**, University of Technology - Vietnam National University.

Honors and Awards

- ACM SIGPLAN Distinguished Paper Award**, PLDI 2014.
- Best Graduate Researcher Award**, UC Davis Computer Science Department, 2014.
- Gold Medal**, ACM Student Research Competition, PLDI 2013.
- PhD Fellowship**, Vietnam Education Foundation, 2009.

Patents

- Sumit Gulwani and Vu Minh Le, **Framework for Data Extraction by Examples**. *US Patent 20,150,254,530*.
- Sumit Gulwani, Ted Hart, Vu Minh Le, Henrique Malvar, Mark Marron, James McCaffrey, Gustavo Araujo Soares, and Benjamin Zorn. **Interactive Data Manipulation Using Examples and Natural Language**. *US Patent 20,150,254,211*.
- Sumit Gulwani, Jonathan Paul de Halleux, Vu Minh Le, Zhendong Su, and Nikolai Tillman. **Generating Program Fragments Using Keywords and Context Information**. *US Patent Application 13/931,625*.

Publications

- ICSE 2016 Chengnian Sun, Vu Le, and Zhendong Su. **Finding and Analyzing Compiler Warning Defects.** *International Conference on Software Engineering*, Austin, TX, May 2016. [ICSE16]
- UIST 2015 Mikaël Mayer, Gustavo Soares, Maxim Grechkin, Vu Le, Mark Marron, Alex Polozov, Rishabh Singh, Ben Zorn, and Sumit Gulwani. **User Interaction Models for Disambiguation in Programming by Example.** *ACM Symposium on User Interface Software and Technology*, Charlotte, NC, November 2015. [UIST15]
- OOPSLA 2015 Vu Le, Chengnian Sun, and Zhendong Su. **Finding Deep Compiler Bugs via Guided Stochastic Program Mutation.** *Systems, Programming, Languages and Applications: Software for Humanity*, Pittsburgh, PA, October 2015. [OOPSLA15]
- ISSTA 2015 Vu Le, Chengnian Sun, and Zhendong Su. **Randomized Stress-Testing of Link-Time Optimizers.** *International Symposium on Software Testing and Analysis*, Baltimore, MD, July 2015. [ISSTA15, 28%]
- PLDI 2014 Vu Le, Mehrdad Afshari, and Zhendong Su. **Compiler Validation via Equivalence Modulo Inputs.** *Programming Languages Design and Implementation*, Edinburgh, UK, June 2014. [PLDI14a, 18%]
🏆 **ACM SIGPLAN Distinguished Paper Award**
- PLDI 2014 Vu Le and Sumit Gulwani. **FlashExtract: A Framework for Data Extraction by Examples.** *Programming Languages Design and Implementation*, Edinburgh, June 2014. [PLDI14b, 18%]
- MobiSys 2013 Vu Le, Sumit Gulwani, and Zhendong Su. **SmartSynth: Synthesizing Smartphone Automation Scripts from Natural Language.** *The International Conference on Mobile Systems, Applications, and Services*, Taipei, June 2013. [MobiSys13a, 16%]
- MobiSys 2013 Vu Le, Jonathan de Halleux, Sumit Gulwani, and Zhendong Su. **Keyword Programming for TouchDevelop.** Video Session, *The International Conference on Mobile Systems, Applications, and Services*, Taipei, June 2013. [MobiSys13b]
- POPL 2013 Earl Barr, Thanh Vo, Vu Le, and Zhendong Su. **Automatic Detection of Floating-Point Exceptions.** *Principles of Programming Languages*, Rome, January 2013. [POPL13, 18%]

Service

- PC Member OOPSLA Artifacts 2013
- External Review Committee PLDI 2016
- External Reviewer FSE 2011, ICSE 2012, ASE 2013, ISSTA 2014, CAV 2015

Research Projects

- PROSE** Microsoft Program Synthesis using Examples (PROSE) SDK is a framework of technologies for programming by examples – automatic generation of programs from input-output examples at runtime. Given a domain-specific language (DSL) and some input-output examples for the desired program's behavior, PROSE algorithms synthesize a ranked set of DSL programs that are consistent with the given examples.
- The PROSE SDK also includes a pre-defined set of specific technologies for various data wrangling domains, such as: FlashFill, a library of string transformations from Microsoft Excel 2013; FlashExtractText, a text extraction library with a DSL of programs that find desired regions in a textual file; FlashExtractWeb, a web extraction library with a DSL of programs that select elements on a web page.
- Critical Software Validation** An umbrella project whose goal is to improve the quality of all aspects of critical software, such as compilers and database engines. We introduce Equivalence Modulo Inputs, a novel methodology to generate valid tests from existing programs to find bugs in compiler optimizers [PLDI14a, OOPSLA15]. We are also developing techniques to find bugs in other important compiler components: link-time optimizers [ISSTA15] and warning infrastructures [ICSE16]. In total, we have reported **400+ new bugs** (majority are miscompilation) in GCC and LLVM, most of which have been confirmed and fixed by developers.
- FlashM** A project that combines various program synthesis techniques and tools (e.g., FlashExtract, FlashFill, and NLyze) to enable users to perform more sophisticated tasks that otherwise cannot be done using a single technique alone [UIST15].
- FlashExtract** A general framework that allows users to extract relevant data from semi-structured documents (text files, web pages, spreadsheets) using examples [PLDI14b]. FlashExtract has been **shipped** as part of the Microsoft Windows Management Framework 5.0 Preview, and Microsoft Azure Operations Management Suites.
- SmartSynth** An end-to-end programming system for synthesizing smartphone automation scripts from natural language descriptions [MobiSys13a]. SmartSynth has been **integrated** into TouchDevelop (www.touchdevelop.com), a popular programming environments on mobile devices [MobiSys14b].
- Ariadne** A practical symbolic execution system that was specifically designed for detecting floating-point exceptions [POPL13]. Ariadne has found 2,091 floating-point exceptions in the GNU Scientific Library (GSL).

Selected Press

Compiler Validation

- Embedded in Academia “... very cool and anyone interested in compiler correctness should read the paper.”
<http://blog.regehr.org/archives/1161>
- Theory and Practice “... a new bug every 2.3 days!”
<http://rgrig.blogspot.com/2014/06/pldi-2014-compiler-validation-via.html>

FlashExtract

- PowerShell Magazine "New kid on the block"
<http://www.powershellmagazine.com/2014/09/09/using-the-convertfrom-string-cmdlet-to-parse-structured-text/>
- LazyWinAdmin "This is super cool!"
<http://www.lazywinadmin.com/2014/09/powershell-convertfrom-string-and.html>
- Development in a Blink "Serious text wrangling"
<http://www.dougfinke.com/blog/index.php/2014/10/25/powershell-convertfrom-string-serious-text-wrangling/>
- Happy SysAdm "...one of the best improvement that came with WMF5.0 and PowerShell v5"
<http://www.happysysadm.com/2014/11/first-look-at-convertfrom-string-in.html>
- Lumagate "I've been waiting for this OMS-feature with anticipation."
<http://www.lumagate.com/news/operations-management-suite-%E2%80%93-custom-fields-/extract-data-feature>

Experience

- 6/2015-present **Research Software Engineer**, Microsoft, Redmond, WA.
Working on the Microsoft Program Synthesis using Examples (PROSE) project.
- 10/2014-06/2015 **Consultant**, Microsoft Research, Redmond, WA.
Worked under the FlashM project that combines multiple synthesis paradigms. Improved FlashExtract logics and ranking to serve other product groups in Microsoft.
- 06/2014-09/2014 **Research Intern**, *Dr. Sumit Gulwani*, Microsoft Research, Redmond, WA.
Worked with the PowerShell team to help transfer FlashExtract to PowerShell (**shipped** as part of the Microsoft Windows Management Framework 5.0 Preview).
- 04/2013-09/2013 **Research Intern** (2 internships), *Dr. Sumit Gulwani*, Microsoft Research, Redmond, WA.
Designed and developed FlashExtract, a general program synthesis framework that enables data extraction from semi-structured documents using examples (**published** in PLDI 2014).
- 06/2012-09/2012 **Research Intern**, *Dr. Jonathan de Halleux and Dr. Sumit Gulwani*, Microsoft Research, Redmond, WA.
Designed and developed a natural language front-end for TouchDevelop to synthesize TouchDevelop scripts from English keywords (**integrated** into TouchDevelop).
- 06/2006-07/2009 **Assistant Lecturer**, University of Technology – Vietnam National University, Ho Chi Minh City, Vietnam.
Co-taught various classes in Programming Languages, Data Structures, and Compiler Construction.

Teaching

- Spring 2010 **Teaching Assistant**, Programming on Parallel Architectures, UC Davis.
Led weekly discussions, graded homework and programming assignments.
- Fall 2008 **Lecturer**, Introduction to Programming, Vietnam National University.
Designed lectures, class projects, and homework assignments.
- Spring 2008 **Assistant Lecturer**, Compiler Construction, Vietnam National University.
Designed class projects, delivered lectures and led discussions.
- Fall 2007 **Teaching Assistant**, Introduction to Programming, Vietnam National University.
Led weekly discussions, graded homework assignments and class projects.
- Spring 2007 **Teaching Assistant**, Compiler Construction, Vietnam National University.
Co-designed class projects, led discussions, and graded homework assignments.

References

Zhendong Su

Professor
Computer Science Department
University of California, Davis
One Shields Ave
Davis CA 95616, USA
✉ su@cs.ucdavis.edu
☎ +1 530 754 5376

Sumit Gulwani

Research Manager & Principal Researcher
Data Platform
Microsoft
One Microsoft Way
Redmond WA 98052, USA
✉ sumitg@microsoft.com
☎ +1 425 706 7709