Software Engineering for Non-Computer Scientists: An Essential Skill for Technical Professionals in the 21st Century

Joseph Hellerstein, PhD

Senior Data Science Fellow in the eScience Institute and Affiliate Professor of Computer Science

David Beck, PhD

Research Assistant Professor of Chemical Engineering and Director of Research – Life Sciences in the eScience Institute

University of Washington, Seattle

Abstract: Computational techniques are essential to work in science and engineering, both in industry and academia. Often, computation artifacts, both code are data, are at the core of a scientific or engineering contribution. As such, there are concerns about the extent to which a computational result can be understood, reproduced, and extended by others.

Today, it is common for non-computer scientists to develop skills in programming, often self-taught. Typically, these skills are sufficient to write modest sized codes for their own purposes. However, this falls far short of what is needed for a 21st century technical career - developing codes and data with colleagues for use by others outside the group. For example, non-computer scientists rarely develop skills with software or data design, coding style, or a knowledge of how to package and distribute software.

This talk discusses our experience with teaching CSE 490 B1, Software Engineering for Biologists. Few of our students had prior experience with programming, although some had written short scripts in MATLAB. The course teaches the basics of programming in MATLAB, python, and SQL as well as skills in object oriented design, testing, and software packaging. We hope to adapt the course to other fields, both in engineering and in social science.

This talk will review dedicated breast PET technology and the requirements for various clinical applications, and report on the development of a combined breast PET / x-ray mammography system (PET/X) under development at the University of Washington. PET/X has rectangular geometry, currently being assessed using simulations, and we are now building detector blocks using Geiger-mode SiPMs. With a focus on therapy monitoring, we seek to minimize the variance of standardized uptake values (SUVs), as SUV-variance will dictate confidence intervals for measuring changes in tracer uptake associated with response to therapy.

Bio: *Joseph L. Hellerstein* is Senior Data Science Fellow in the eScience Institute and Affiliate Professor of Computer Science, both at the University of Washington, Seattle, Washington. Previously, Dr. Hellerstein managed the Computational Discovery Department at Google (2008-2014), was a Principal Architect at Microsoft Corp. in Redmond, WA (2006 to 2008), and

founded/directed the Adaptive Systems Department at the IBM Thomas J. Watson Research Center in Hawthorne, NY (1984 to 2006). Dr. Hellerstein received the PhD in computer science from the University of California at Los Angeles. He has published approximately 200 peer-reviewed papers, 30 patents, and two books. He has taught at Columbia University and the University of Washington, and has served on numerous program committees and government advisory panels. Dr. Hellerstein is a recipient of the 2007 IEEE/IFIP Stokesberry Award, and is a Fellow of the IEEE.