Computer Science Colloquium

"Data Mining- The Missing Link in Empirical Software Engineering"

Professor Tim Menzies

Associate Professor
West Virginia University

Abstract

There is an old joke where two scientists are staring at a blackboard. In between the equations on the left and the equations on the right, it is written "and then a miracle occurs". The older scientist points to this missing link and comments "I think you need to be more explicit in step two"
(http://socsci2.ucsd.edu/~aronatas/project/cartoon.math.gif)

If you read the empirical SE literature, you can see a similar "missing link" in recent descriptions of how to conduct empirical SE studies. In one paper I just read, there was:

• 9 pages: on selecting methods
• 3 pages: on research questions
• 2 pages: on different forms of "empirical truth"
• 2 pages: on empirical validity
• 1 page: on the role of theory building
• 1 page: on data collection techniques
• And 0 pages: on data analysis (and then a miracle happens)

Speaking as a data mining researcher, I'm asserting that selecting data analysis methods deserves more than 0 pages. Data analysis is not a simple matter of loading up a box with data, then pressing a button to produce the output. The actual procedure is far more complex and, if done correctly, can lead to radical new insights that drive the next generation of the theorizing and experimentation. This talk will illustrate this process with detailed examples from effort estimation and defect prediction.

Biography

Dr. Tim Menzies (tim@menzies.us) has been working on advanced modeling and AI since 1986. He received his PhD from the University of New South Wales, Sydney, Australia and is the author of over 164 referred papers. A former research chair for NASA, Dr. Menzies is now an associate professor at the West Virginia University's Lane Department of Computer Science and Electrical Engineering. For more information, visit his web page at http://menzies.us.

The colloquium is open to the general public

Date: Friday, October 23, 2009
Time: 10:30am – 11:30am
Location: Room: ECSS 2.410

(Refreshments will be served at 10:15am)

The University of Texas at Dallas
ENGINEERING/COMPUTER SCIENCE BLDG. EC 31, P.O. BOX 830688, RICHARDSON, TEXAS 75083-0688 (972) 883-2190