

On The Naturalness of Software

Prem Devanbu

Department of Computer Science
University California
Davis, CA, USA

ABSTRACT

Programming languages, like their "natural" counterparts, are rich, powerful and expressive. But while skilled writers like Zadie Smith, Aravind Adiga, and Salman Rushdie delight us with their elegant, creative deployment of the power and beauty of English, most of what us regular mortals say and write everyday is Very Repetitive and Highly Predictable.

This predictability, as most of us have learned by now, is at the heart of the modern statistical revolution in speech recognition, natural language translation, question-answering, etc. We will argue that in fact, despite the power and expressiveness of programming languages, most <<Software>> in fact is <<also>> quite repetitive and predictable, and can be fruitfully modeled using the same types of statistical models used in natural language processing. We present some practical applications of this rather unexpected finding, and present a research vision arguing that this phenomenon is potentially rich in both scientific questions, and engineering promise.

This is an international effort is currently funded by the U.S NSF and the UK EPSRC. Active collaborators include Zhendong Su at UC Davis, Roni Rosenfeld and William Cohen of CMU, Earl Barr and Mark Harmon of University College, London, Mark Gabel of UT Dallas, and Charles Sutton of the University of Edinburgh. There is a lot to do, and we welcome more collaborators.

BIOGRAPHY

Prem Devanbu received his B.Tech from the Indian Institute of Technology in Chennai, India, before you were born, and his PhD from Rutgers in 1994. After spending nearly 20 years as both a developer and a researcher at AT&T Bell Labs and its various offshoots, he left Industry to join the CS faculty at UC Davis in late 1997. He serves on the editorial boards of the Empirical Software Engineering and the Wiley Journal of Software Process and Maintenance. He was PC Chair of ICSE in 2010 and SIGSOFT FSE in 2006, and ICSR in 1998. He has previously served on the editorial boards of both the IEEE and ACM Transactions journals relating to software engineering. He has published a few papers, given a few keynotes, bagged a few patents and won a few awards over the years; he even has his own web page. His research interest is mainly in exploiting Good Data to help Software Engineers live better, longer, happier, more meaningful lives.