A Study of String Hashing in Programming Languages and Libraries

CS Colloquium

Speaker: Clinton Jeffery and Robert Heckendorn

Date & Time: Nov 18, 2013, 3:30 – 4:20 PM

Place: Engineering Physics Building, EP 205

Abstract

String hashing is heavily used; it is almost everywhere in modern software systems that depend on hash tables or dictionaries or maps. This talk will illustrate how even fairly clever hash functions can have flaws that result in poor performance on some data sets. A survey of widely used hash functions in several popular programming languages and libraries will be presented, along with a real world case study in which Unicon's string hashes were tuned recently, resulting in a dramatic performance gain in a customer application at the National Library of Medicine (NIH).

Bios

Dr. Clinton Jeffery completed his Ph.D. in CS under Ralph Griswold at the University of Arizona in 1993. His doctoral research developed a framework for program execution monitoring, with a substantial subcomponent examining very high level language support for graphics and program visualization. As a faculty member at the University of Nevada Las Vegas, he developed the Unicon programming language with support from the National Library of Medicine at NIH, a project which has continued to this day. At the University of Idaho he and his students work on automatic debugging and was fortunate to obtain NSF support for a project in the area of educational collaborative virtual environments.

Dr. Robert Heckendorn did his Ph.D. in CS at Colorado State University in 1999, on patterns of epistasis in evolutionary algorithms. His current research is in evolutionary computing, including: game theory and international conflict, the evolution of social organization, epistasis in artificial and natural worlds, and COTSbots: inexpensive but computationally powerful robots. The latter consist of commercial radio-controlled vehicles fitted with smart phones to serve as eyes, ears, motion sensors, and brains, for exploring complex behaviors and machine learning.