Dr. Erik Ferragut

Title:  Nonparametric Bayesian Modeling for Automated Database Schema Matching

Abstract:

The problem of merging databases arises in many government and commercial applications. Schema matching, a common first step, identifies equivalent fields between databases. A new schema matching framework is introduced for building nonparametric Bayesian models for each field and comparing them by computing the probability that a single model could have generated both fields. The speaker will introduce the nonparametric Bayesian paradigm and motivate its use. Experiments show that the framework is more accurate and faster than the existing instance-based matching algorithms in part because of the use of nonparametric Bayesian models. The talk presents joint work with Jason Laska (ORNL).

BIO:

Dr. Erik M. Ferragut is a Cyber Security Research Scientist and an Applied Research Mathematician in the Cyber and Information Security Research Group within the Oak Ridge National Laboratory. His research interest is primarily in applying probabilistic modeling and other mathematical methods to diverse problems. Recent work has including analysis of cyber-physical system security and developing scalable machine learning solutions for data fusion and fraud detection. Previously, he worked as a cryptologic researcher at NSA for over 10 years. He earned his Ph.D. in Mathematics from the University of Michigan, Ann Arbor in 2003.