

# HCI Meets Data Mining: Principles and Tools for Big Data Analytics

Duen Horng (Polo) Chau, College of Computing, Georgia Tech

June 13, 10:00 – 12:00  
CoRE Building 431

Hosted by James Abello

## Abstract

Learn key data mining concepts and tools, and how to integrate them with HCI methods for big data analytics, through non-technical introduction and case studies, and technical examples. This two-part course takes a practical approach to introduce you to the principles, tools and pitfalls in big data analytics.



**Part 1:** A non-technical introduction illustrating where HCI and data mining as fields of research and practice can benefit from each other with illustrative case studies, followed by a review of tools for analyzing datasets from small to huge.

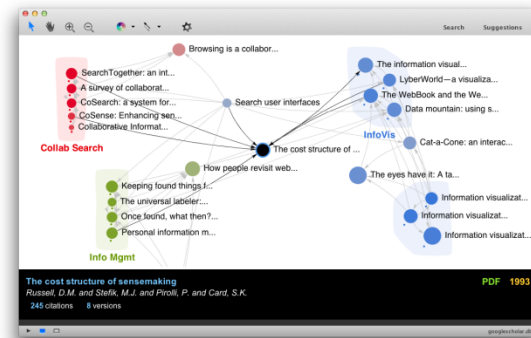
**Part 2:** A more technical discussion of how to "do it right", such as: How to choose a "big data" platform for your work (or do you need one at all)? How to find an algorithm that is right for your data? How to evaluate your approach appropriately? And more...

## Features:

- Learn the foundation of data mining and machine learning methods
- Learn how to integrate these techniques into your research
- Learn about practical tools for storing and analyzing large datasets with commodity PCs
- Understand common pitfalls in applying machine learning methods, and how to avoid them, e.g., feature selection, choosing the right algorithms
- Be inspired by exciting research problems and directions in bridging data mining and HCI

## Speaker

Prof. Polo Chau is an assistant professor at Georgia Tech's College of Computing, and the Associate Director of MS in Analytics program. Polo holds a Ph.D. in Machine Learning and a Masters in HCI, both from Carnegie Mellon, and has been working at the intersection of HCI and data mining since 2004, to create scalable, interactive tools for big data analytics. Polo's thesis on bridging HCI and data mining for making sense of large network data won received Carnegie Mellon's Distinguished Computer Science Dissertation Award, Honorable Mention. He teaches the "Data and Visual Analytics" course at Georgia Tech. Polo is the only two-time Symantec fellow. He contributes to the PEGASUS peta-scale graph mining that won an Open Source Software World Challenge Silver Award. Polo's NetProbe auction fraud detection research appeared on The Wall Street Journal, CNN, TV and radio. His Polonium malware detection technology protects 120 million people worldwide. Polo designed Carnegie Mellon's ID card.



For questions or more information, please contact Dr. Eugene Fiorini, Associate Director of DIMACS and Program Coordinator ([gfiorini@dimacs.rutgers.edu](mailto:gfiorini@dimacs.rutgers.edu)). The Workshop is organized by the DIMACS REU program (<http://dimacs.rutgers.edu/REU/>).

