The Department of Health Policy and Management
Population Informatics Lab Seminar

"LINKING RECORDS: CHALLENGES AND SOLUTIONS"

Luiza Antonie, Ph. D.
Associate Professor
School of Computer Science
University of Guelph, Canada

Tuesday, April 9, 2019

SPH Classroom Building

Room 147

4:00pm – 5:00pm

RSVP to sulkipark828@tamu.edu

Biography

Dr. Luiza Antonie is an assistant professor in the School of Computer Science at the University of Guelph, Canada. She is the lead computer scientist of the social science-based People in Motion project which links people across historical Canadian censuses. She received her PhD in Computing Science from University of Alberta with specialization in data mining. Her research interests include applied record linkage, associative classifiers and health informatics. She has published articles on these topics in top venues. She serves as a reviewer and member of program committee in journals and conferences related to data mining. She was a co-organizer for the Workshop for Women in Machine Learning in 2008, one of the organizing chairs for Ontario Celebration of Women in Computing in 2014, co-chair for the Graduate Student Symposium at the Canadian Conference on AI in 2018 and a co-organizer for the 1st, 2nd, 3rd and 4th Data Integration and Applications workshops.

Abstract

Linking multiple databases to create longitudinal data is an important research problem with multiple applications. Longitudinal data allows analysts to perform studies that would be unfeasible otherwise. In this talk, I introduce record linkage and discuss a system we designed to link census databases in order to create longitudinal data that allow tracking people over time. The goal of the linking is to identify the same person in multiple census collections. Data imprecision and the lack of unique personal identifiers make this task a challenging one. We design and employ a record linkage system that incorporates a supervised learning module for classifying pairs of records as matches and non-matches. We show that our system performs large scale linkage producing high quality links and generating sufficient longitudinal data to allow meaningful population studies. These longitudinal data have already been used to investigate trends and to address questions about society, history and economy, and this comparative, systematic research would not be possible without the linked data.