Introduction to the Minitrack on Text Analytics

Evangelos Milios
Faculty of Computer Science
Dalhousie University
Halifax, Nova Scotia
eem@cs.dal.ca

Carolyn Watters
Faculty of Computer Science
Dalhousie University
Halifax, Nova Scotia
cwatters@dal.ca

This minitrack is focused on systems and tools that enable people to engage in meaningful ways with the information encapsulated in massive textual datasets. The papers and the discussions center on the extraction and visualization of concepts and relationships, real-time interaction with text datasets including novel visualization and interaction techniques for users.

Text analytics is a core technology for a range of text based data contexts from social media to large operational management decision support systems. The techniques explored include the extraction of concepts and relationships from large noisy text corpora or textual components of large corpora. Equally importantly, text analytics provides core technology and insight into the models and interactions that support people in understanding the information extracted.

The papers in this first year of the Text Analytics Minitrack includes two papers with focus on extraction of information and two papers on sensemaking of that information . The first paper, *Automating Discovery of Dominance in Synchronous Computer-Mediated Communication,* explores the detection of behavioral patterns in chat transcripts. The second paper, *Large-scale Web Page Classification*, addresses content-based text classification when the number of classes is in the tens or hundreds of thousands and classes are highly unbalanced in the training data. The final paper, *Mapping the Territories of Privacy. Textual Analysis of Privacy Frameworks in American Mainstream News,* provides an interdisciplinary approach by exploring the detection of privacy frames in American media. *Word Cloud Explorer: Text Analytics Based on Word Clouds* explores the usefulness of word clouds extended with additional information and interactivity in text analytics.