Workshop Schedule

Presenters are listed in italic.

**Introduction**
8:50-9:00am  
*Michael W. Berry*, University of Tennessee, Knoxville

**Keynote**
9:00-10:00am  
“Visualization for Text Analysis”  
*John R. Goodall*, Oak Ridge National Laboratory

**Abstract:**

Human-generated, typically unstructured, content has exploded due to the popularity and availability of Internet-based technologies: social networking sites, blogs and microblogs, shared documents, and ever increasing reliance on email, instant messaging, and SMS. These represent treasure troves of potentially useful data to reveal patterns and trends, to understand how people foster relationships and how they use language to communicate. Text mining promises to expose interesting characteristics, hidden patterns and key relationships within these textual corpora. In order to make the results of textual mining and analytics understandable to humans, these algorithms would benefit from integration with information visualization techniques. This talk will address some of the research and practice in information visualization that may be useful in interactions with textual data sets, as well as in exploring the results of text mining algorithms that expose important concepts and relationships.

10:00-10:30am  
Coffee Break

**Session I: Matrix Factorization and Clustering Models** (Michael W. Berry, Chair)

10:30-11:00am  
“A Matrix Factorization Framework for Jointly Analyzing Multiple Nonnegative Data Sources”  
*Sunil Gupta, Dinh Phung, Brett Adams, and Svetha Venkatesh*

11:00-11:30am  
“Learning Hidden Markov Models Using Probabilistic Matrix Factorization”  
*Ashutosh Tewari* and *Michael Giering*

11:30am-12:00pm  
“Mining Research Cycles with Adapted Hierarchical Clustering”  
*Dan He* and *Douglas S. Parker*
12:00-1:30pm  Lunch

**Session II:**  **Applications in Biology and Social Networking** (Jacob Kogan, Chair)

1:30-2:00pm  “Nonnegative Tensor Factorization of Biomedical Literature for Analysis of Genomic Data”
*Sujoy Roy*, Ramin Homayouni, Michael Berry, and Andrey Puretskiy

2:00-2:30pm  “A Weighted Density-based Approach for Identifying Standardized Items that are Significantly Related to the Biological Literature”
*Omar Al-Azzam*, Jianfei Wu, Loai Al-Nimer, Charith Chitraranjan, and Anne Denton

2:30-3:00pm  “Scam Detection in Twitter”
*Xiaoling Chen*, Rajaratnam Chandramouli, and Koduvayur Subbalakshmi

3:00-3:30pm  Coffee Break

**Session III:**  **Dimension Reduction** (Michael W. Berry, Chair)

3:30-4:00pm  “Dimensionality Reduction for Information Retrieval using Vector Replacement of Rare Terms,”
*Tobias Berka* and Marian Vajtersic

4:00pm  Adjourn