# Workshop Schedule

*Presenters are listed in *italic.*

## Introduction
8:00-8:10am  
*Michael W. Berry,* University of Tennessee, Knoxville

## Keynote
8:10-9:00am  
“Using Advanced Mathematical Techniques to Help Bring Electronic Discovery Under Control”  
*John Tredennick* and *Bruce Kiefer,* Catalyst Repository Systems, Inc.

**Abstract:**

With digital content exploding, corporations are seeking new and better ways to tame the eDiscovery dragon and reduce the cost of litigation review. In the past, legal teams would divide up the case documents and put eyeballs on every page. With million-document cases now becoming almost routine, clients are balking at the costs. Rather than pay expensive hourly rates to look at documents, they are turning to advanced mathematics and statistical techniques to reduce review populations. In this talk, we will explore real-world techniques used by Catalyst Repository Systems, one of the leading international document hosting companies for complex legal matters. Catalyst's founder and CEO, John Tredennick, a longtime trial lawyer handling these kinds of big cases, will talk about the problem space and how traditional methods for dealing with eDiscovery have failed. Bruce Kiefer, Catalyst's VP of Operations, will talk about how we use advanced clustering and search-based analytics techniques derived from non-negative matrix factorization techniques to help address the problem.

## Session I: Concept Tagging

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Presenters</th>
</tr>
</thead>
</table>
| 9:00-9:30am| “Discovering Generalized Concepts from Documents Using a Category Graph”  
*Thomas Vacek,* John Joseph, Daniel Boley |
| 9:30-10:00am| “A Hybrid Recommender System Based on Weighted Tags”  
*Huizhi Liang,* Yue Xu, Yuefeng Li, Richi Nayak |

10:00-10:30am Coffee Break
Session II:  
**Applied Linear Algebra**

10:30-11:00am  
“A Fast Algorithm for Nonnegative Tensor Factorization using Block Coordinate Descent and Active Set-like Method”  
Krishnakumar Balasubramanian, Jingu Kim, Andrey Puretskiy, Michael W. Berry, Haesun Park

11:00-11:30am  
“Orthogonal Nonnegative Matrix Factorization for Multi-type Relational Clustering”  
Chengcheng Shen, Ying Liu

11:30-12:00am  
“EDLSI with PSVD Updating”  
April Kontostathis, Erin Moulding, Raymond Spiteri

12:00-1:15pm  
Lunch

Session III:  
**Document Clustering and Large-Scale Text Mining**

1:30-2:00pm  
“Large-Scale Text Mining”  
Alan Ratner

2:00-2:30pm  
“Improving Document Clustering Using a Hierarchical Ontology Extracted From Wikipedia”  
Mostafa Hassan, Fakhreddine Karray, Mohamed S. Kamel

2:30-3:00pm  
“Regularized Co-Clustering on Manifold”  
Chengcheng Shen, Ying Liu

3:00-3:30pm  
Coffee Break

3:30-4:00pm  
“Enhancing Document Clustering Using Hybrid Models for Semantic Similarity”  
Ahmed Farahat, Mohamed Kamel

4:00-4:30pm  
“Information Bottleneck Co-Clustering”  
Pu Wang, Carlotta Domeniconi, Kathryn Laskey