

The logo for Text Mining 2006 is centered on a light yellow square background. The word "TEXT" is in a bold, black, sans-serif font, with the letter "E" in orange and slightly offset to the left. Below "TEXT" is the word "MINING" in a bold, black, sans-serif font. At the bottom is the year "2006" in a large, bold, orange, sans-serif font.

**TEXT  
MINING  
2006**

Organizers: Michael W. Berry (Tennessee)  
Malu Castellanos (Hewlett-Packard)

Sixth SIAM International Conference on Data Mining, Bethesda, MD

April 22, 2006

# Author Demographics and Acceptance Rate

- ▶ 54 authors from industry, academia, and research laboratories
- ▶ Authors from 13 different countries submitted 25 papers
- ▶ 10 papers and 4 posters selected
- ▶ No distinction made for student-authored papers in review process
- ▶ Submissions from past workshops: 2001 (13), 2002 (11), 2003 (16)

# Acknowledgments

- ▶ Special thanks to Murray Browne for his help with preparing the proceedings volume and Jeff Romaniuk for creating the logo and workshop poster.
- ▶ Thanks to the Program Committee for their diligent efforts in reviewing the papers and posters.
- ▶ We are grateful to SAS and PureDiscovery for sponsoring the workshop and providing funds to support planning, materials, and student travel and registrations.



Superior software that gives  
you *The Power to Know.*



**TEXT  
MINING  
2006**

## Morning Schedule - Part I

- ▶ 8:15-8:20am – **Introduction** (M.W. Berry, M. Castellanos)
- ▶ 8:20-9:00am – **Keynote:** *The Needle in the Haystack Problem: Discovering Anomalies in Text Documents* (Ashok Srivastava, NASA Ames)
- ▶ 9:00-10:00am – **Session I:** Topic Dection and Tracking  
*Confirming Protein-Protein Interactions by Text Mining* (D. Otasek, Ontario Cancer Institute)

*Entropy Based Measure Functions for Analyzing Time Stamped Documents* (P. Chundi, University of Nebraska at Omaha)

- ▶ 10:00-10:30am – Coffee Break

**TEXT  
MINING  
2006**

## Morning Schedule - Part II

▶ 10:30-11:30am – **Session II: Text Classification**

*One-Sided Non-Negative Matrix Factorization and Non-Negative Centroid Dimension Reduction for Text Classification* (H. Park, Georgia Tech)

*Document Author Classification using Generalized Discriminant Analysis* (T. Moon, Utah State University)

▶ 11:30-12:15pm – **Session III: Poster Presentations**

*Automatically Adjusting Content Taxonomies for Hierarchical Classification* (J. Zhang, America Online Inc.)

*Using Query History to Prune Query Results* (D. Waegel, Ursinus College)

*OPTICS on Text Data: Experiments and Test Results* (S. Roy, IBM India Research Lab)

*ZIP and Data Document Visualization* (D. Alvarez-Medina, CICESE, Mexico)

**TEXT  
MINING  
2006**

## Afternoon Schedule

▶ 1:45-3:15pm – **Session IV: Clustering Algorithms**

*Ontology-based Distance Measure for Text Clustering* (L. Jing, University of Hong Kong)

*Model-based Overlapping Co-Clustering* (M. Shafiei, Dalhousie University)

*Scaling Clustering Algorithms with Bregman Distances* (J. Kogan, University of Maryland, Baltimore County)

▶ 3:15-3:45pm – Coffee Break

▶ 3:45-4:45pm – **Session V: LSA and Vector Space Models**

*Latent Semantic Analysis and Fiedler Embeddings* (B. Hendrickson, Sandia National Labs)

*Exploring Term-Document Matrices from Matrix Models in Text Mining* (E. Gallopoulos, University of Patras)

TEXT  
MINING  
2006