System logs contain a wide variety of information about system status and health, including events from various applications, daemons and drivers, as well as sampled information such as resource utilization statistics. As such, these logs represent a rich source of information for the analysis and diagnosis of system problems and prediction of future system events. However, their lack of organization and the general lack of semantic consistency between information from various software and hardware vendors means that most of this information content is wasted. Indeed, today's most popular log analysis technique is to use regular expressions to either detect events of interest or to filter the log so that a human operator can examine it manually. Clearly, this captures only a fraction of the information available in these logs and does not scale to the large systems common in business and supercomputing environments.

This workshop will focus on novel techniques for extracting operationally useful information from existing logs and methods to improve the information content of future logs. Topics include but are not limited to:

- Reports on publicly available sources of sample log data.
- Log anonymization
- Log feature detection and extraction
- Prediction of malfunction or misuse based on log data
- Statistical techniques to characterize log data
- Applications of Natural-Language Processing (NLP) to logs
- Scalable log compression
- Log comparison techniques
- Methods to enhance and standardize log semantics
- System diagnostic techniques
- Log visualization
- Analysis of services (problem ticket) logs
- Applications of log analysis to system administration

Papers limited to 6 2-column pages using >=10pt font.

Workshop Chair: Greg Bronevetsky (Lawrence Livermore National Laboratory)

mailto:mgreg@bronevetsky.com