A Software Infrastructure for Research in Textual Data Mining by Dr. William M. Pottenger & Lars Holzman

Poster #18 Abstract: Few tools exist that address the challenges facing researchers in the Textual Data Mining (TDM) field. Some are too specific to their application, or are prototypes not suitable for general use. More general tools often are not capable of processing large volumes of data. We have created a Text Mining Infrastructure (TMI) that incorporates both existing and new capabilities in a reusable framework conducive to developing new tools and components. TMI adheres to strict guidelines that allow it to run in a wide range of processing environments – as a result, it accommodates the volume of computing and diversity of research occurring in TDM. A unique capability of TMI is support for optimization. This facilitates text mining research by automating the search for optimal parameters in text mining algorithms. In this poster we describe the use of TMI. We also discuss how the TMI utilizes existing machine-learning libraries, thereby enabling researchers to continue and extend their endeavors with minimal effort. Towards that end, TMI is available on the web at hddi.cse.lehigh.edu.

Biographical Information: **Dr. William M. Pottenger** (please see Poster #17). **Lars Holzman** is currently a Senior at Lehigh University. He is working towards a B.S. in Computer Science with minors in Cognitive Science and Philosophy. While at Lehigh University he has worked with Professor William M. Pottenger in the Parallel and Distributed Textual Data Mining lab. He has co-authored a number of papers and will be presenting at the 15th IEEE International Conference on Tools with Artificial Intelligence in November. After completing his bachelors he will be pursuing a Ph.D. in Computer Science with a focus on Artificial Intelligence. His eventual goal is to become tenure track faculty member at a research institution.