B15 Application of Natural Language Processing to the Digital Forensic Process

Mark Pollitt, MS, University of Central Florida, PO Box 162367, Orlando, FL 32816-2367; and Anne Diekema, PhD, Center for Natural Language Processing, Syracuse University, 335 Hinds Hall, Syracuse, NY 13244*

After attending this presentation, attendees will understand the level structure of natural language processing, the corresponding levels of abstraction and access in digital forensics, and how the two taxonomies are related.

This presentation will impact the forensic community by introducing the digital forensic community to the theories and approaches utilized in the natural language processing community.

Since the second World War, computer scientists and others have struggled to find computational methods to translate, understand, and mimic human communications. What has evolved is an interdisciplinary approach known collectively as natural language processing (NLP). The literature in this community includes the notion of "levels of language" which describe the different ways in which "text" in the broadest sense communicates meaning. These include: phonology, morphology, lexical, syntactic, semantic, discourse, and pragmatic.

The digital forensic community faces a somewhat similar problem in that meaning is stored on computers at a number of different levels. The context and therefore the meaning of any particular data, from a digital forensic perspective, can be altered by the various levels of access/abstraction including the physical media, operating system, file system, application, and content. The metadata from each of these layers provide additional context that shapes the meaning of the data.

This presentation will provide a brief overview of the history of NLP, an explanation of the NLP levels of language, a review of the digital forensic levels of access/abstraction, discuss the similarities of these two processes and map their correspondence with the goal of identifying NLP techniques and methodologies that can be applied to digital forensics.

Natural Language Processing, Digital Evidence, Digital Forensics

 ** Presenting Author*