Going beyond Identifinder

Extraction of common named entities (Named Entity Recognition, or NER) from normal English documents is largely a solved problem. As part of an ongoing research project here, we are attempting to go beyond standard NER in several dimensions. We have investigated ensemble approaches to improving NER quality, rapid development of NER in other languages, and NER co-reference both within and across documents. We have also made an initial investigation into NER in short informal messages (Twitter "tweets"). I will describe the context of this work, our approaches and results so far, and then mention future directions that we are considering. This talk presents joint work with Anatole Gershman, Rushin Shah, and Bo Lin.

Robert Frederking is a Senior Systems Scientist in the Language Technologies Institute (LTI) at Carnegie Mellon, where he is also SCS Associate Dean for Graduate Education, and the Chair of the LTI's graduate programs. His research in information extraction began with the Radar project. His research interests also include machine translation and other areas of language technologies. He received his PhD in Computer Science/Artificial Intelligence from Carnegie Mellon in 1986. He received his BS in Computer Engineering from CWRU in 1977, and worked on CWRUNET, an early computer network. He has also worked at Carnegie Group Inc., the Robotics Institute, and Siemens Research in Munich, Germany.

* Find more information about the talk and LTI colloquium at:
  http://www.cs.cmu.edu/afs/cs.cmu.edu/project/cmt-55/lti/Courses/700/