

The Knowledge Base Population Task: Challenges for Information Extraction (invited talk)

Ralph Grishman
New York University
grishman@cs.nyu.edu

Abstract

The Knowledge Base Population (KBP) task, being run for the past 3 years by the U.S. National Institute of Standards and Technology, is the latest in a series of multi-site evaluations of information extraction, following in the tradition of MUC and ACE. We examine the structure of KBP, emphasizing the basic shift from sentence-by-sentence and document-by-document evaluation to corpus-based extraction and the challenges it raises for cross-sentence and cross-document processing. We consider the problems raised by the limited amount and incompleteness of the training data, and how this has been (partly) addressed through such methods as semi-supervised learning and distant supervision. We describe some of the optional tasks which have been included – rapid task adaptation (last year), temporal analysis (this year), cross-lingual extraction (planned for next year) – and others which have been suggested.