

Toward Large-Scale Shallow Semantics for Higher-Quality NLP

Eduard Hovy

Information Sciences Institute
University of Southern California

<http://www.isi.edu/natural-language/people/hovy/bio.html>

Abstract. Building on the successes of the past decade's work on statistical methods, there are signs that continued quality improvement for QA, summarization, information extraction, and possibly even machine translation require more-elaborate and possibly even (shallow) semantic representations of text meaning. But how can one define a large-scale shallow semantic representation system and contents adequate for NLP applications, and how can one create the corpus of shallow semantic representation structures that would be required to train machine learning algorithms? This talk addresses the components required (including a symbol definition ontology and a corpus of (shallow) meaning representations) and the resources and methods one needs to build them (including existing ontologies, human annotation procedures, and a verification methodology). To illustrate these aspects, several existing and recent projects and applicable resources are described, and a research programme for the near future is outlined. Should NLP be willing to face this challenge, we may in the not-too-distant future find ourselves working with a whole new order of knowledge, namely (shallow) and doing so in increasing collaboration (after a 40-years separation) with specialists from the Knowledge Representation and reasoning community.