

Modeling Unrestricted Coreference in OntoNotes

The importance of coreference resolution for the entity/event detection task, namely identifying all mentions of entities and events in text and clustering them into equivalence classes, has been well recognized in the natural language processing community. Automatic identification of coreferring entities and events in text has been an uphill battle for several decades, partly because it can require world knowledge which is not well-defined and partly owing to the lack of substantial annotated data.

The OntoNotes project (<http://www.bbn.com/ontonotes/>) — a collaborative effort between BBN Technologies, University of Colorado, University of Southern California (ISI), University of Pennsylvania and Brandeis University — has created a large-scale, accurate corpus for general anaphoric coreference that covers entities and events not limited to noun phrases or a limited set of entity types. The Linguistic Data Consortium (LDC) has agreed to make it freely available to the research community. The coreference layer in OntoNotes constitutes one part of a multi-layer, integrated annotation of shallow semantic structure in text with high inter-annotator agreement. In addition to coreference, this data is also tagged with syntactic trees, high coverage verb and some noun propositions, partial verb and noun word senses, and 18 name entity types.

Modeling unrestricted coreference in the OntoNotes data is the shared task for this year. More information about the task can be found in the [Task Description](#) tab.

Organizers

[Sameer Pradhan](#) (Chair) Raytheon BBN Technologies, Cambridge, MA [Mitchell Marcus](#) ,
University of Pennsylvania, Philadelphia, PA
[Martha Palmer](#)
, University of Colorado, Boulder, CO Lance Ramshaw, Raytheon BBN Technologies,
Cambridge, MA Ralph Weischedel, Raytheon BBN Technologies, Cambridge, MA
[Nianwen Xue](#)
, Brandeis University, Waltham, MA