

Stanford CoreNLP NER (Named Entity Recognition) Annotator

The CoNLL table, produced in output (in csv format) by the Stanford CoreNLP parsers (Klein and Manning 2003), contains a field NER (Named Entity Recognition). The NER can also be obtained as an independent annotator without running the full parser.

For English, the Stanford CoreNLP, by default through the NERClassifierCombiner annotator, recognizes the following NER values: named (PERSON, LOCATION, ORGANIZATION, MISC), numerical (MONEY, NUMBER, ORDINAL, PERCENT), and temporal (DATE, TIME, DURATION, SET) entities (12 classes). By adding the regexner annotator and using the supplied RegexNER pattern files adds support for the fine-grained and additional entity classes EMAIL, URL, CITY, STATE_OR_PROVINCE, COUNTRY, NATIONALITY, RELIGION, (job) TITLE, IDEOLOGY, CRIMINAL_CHARGE, CAUSE_OF_DEATH (11 classes) for a total of 23 classes. **All NLP Suite algorithms that use the NER annotator export the full set of NER classes.**

References

Klein, Dan and Christopher D. Manning. 2003. "Accurate Unlexicalized Parsing." In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, pp. 423–430. Association for Computational Linguistics.