

Frequently Asked Questions

- **Can I use the development data to train the system for the final test run?**

Answer: Yes. It is fine to use the annotation on the development data as additional training material for the final evaluated system.

- **We are getting a UnicodeDecodeError when we try to create the *_conll files**

```
python skeleton2conll.py ../../[...]/wb/a2e/00/a2e_0002.parse
../v1/[...]/wb/a2e/00/a2e_0002.v1_auto_skel -edited --text      Traceback (most recent call last):
  File "skeleton2conll.py", line 737, in      start(input_fname, conll_fname, output_fname,
encoding, changes)  File "skeleton2conll.py", line 674, in start      rows.append("
".join(columns)) UnicodeDecodeError: 'ascii' codec can't decode byte 0xd8 in position 1: ordinal
not in range(128)
```

Answer:

This was happening owing to a bug in the skeleton2conll.py script. Please download the new version of the scripts and that should fix the issue.

- **Each *_conll file is split into one or many “parts”. Do we consider each part as a separate document?** >

Answer: Yes. As mentioned in the OntoNotes release document, some files were too long for a complete coreference annotation, so we had to break them into smaller pieces. Therefore, each part behaves as a separate document. We have, in some cases, tried to merge multiple contiguous parts into one to form longer, coherent parts, and given more time and feasibility we plan to merge as many as possible in later releases.

- **Do the participants have to perform mention detection and coreference resolution, or will the mention bracketing be available in the test sets?**

Answer: Systems are expected to perform **both** mention detection **and** coreference resolution. The final coreference column in the test set will be blank (i.e., a series of “-”).

- **Are singleton mentions not tagged in OntoNotes? Should we filter them out before scoring ourselves, and before submitting the final test run?**

Answer: Yes, by design, we have only tagged mentions that have one or more other coreferent mentions in the document. Therefore, if your system produces singleton mentions, they should be filtered out before scoring or submitting the final runs. The scorer is not designed to filter them out implicitly, and would consider them to be spurious mentions.

- **There are two versions of the CEAF metric: i) Mention based CEAF, or CEAFM; and**

ii) Entity based CEAF, or CEAFE. Which of these will be used for the final average score?

Answer: We will use the Entity based (CEAFE) metric and compute the final score based on the following formula using the F-scores of each metric:

$$(MUC + BCUBED + CEAFE)/3$$

- **Can we submit the final test results in one big file?**

Answer: Yes, you may do that, as long as all the document metadata — including the part numbers are preserved in the right format.

- **What is the citation for the shared task introduction paper?**

Answer: Please use the following bibtex citation:

```
@inproceedings{pradhan-etal-conll-st-2011-ontonotes, author = {Sameer Pradhan and Lance Ramshaw and Mitchell Marcus and Martha Palmer and Ralph Weischedel and Nianwen Xue}, title = {CoNLL-2011 Shared Task: Modeling Unrestricted Coreference in OntoNotes}, booktitle = {Proceedings of the Fifteenth Conference on Computational Natural Language Learning (CoNLL 2011)}, month = {June}, year = {2011}, address = {Portland, Oregon} }
```