CALL FOR PARTICIPATION
CoNLL 2017 Shared Task:
Multilingual Parsing from Raw Text to Universal Dependencies

Website: http://universaldependencies.org/conll17

Email contact: ud.conll.shared.task.2017@gmail.com

Ten years ago, two CoNLL shared tasks were a major milestone for parsing research in general and dependency parsing in particular. For the first time dependency treebanks in more than ten languages were available for learning parsers; many of them were used in follow-up work, evaluating parsers on multiple languages became a standard; and multiple state-of-the art, open-source parsers became available, facilitating production of dependency structures to be used in downstream applications. While the 2006 & 2007 tasks were extremely important in setting the scene for the following years, there were also limitations that complicated application of their results: 1. gold-standard tokenization and tags in the test data moved the tasks away from real-world scenarios, and 2. incompatible annotation schemes made cross-linguistic comparison impossible. CoNLL 2017 will pick up the threads of the pioneering tasks and address the two issues just mentioned.

The focus of the 2017 task is learning syntactic dependency parsers that can work in a real-world setting, starting from raw text, and that can work over many typologically different languages, even surprise languages for which there is little or no training data, by exploiting a common syntactic annotation standard. This task has been made possible by the Universal Dependencies initiative (UD, http://universaldependencies.org/), which has developed treebanks for 40+ languages with cross-linguistically consistent annotation and recoverability of the original raw texts. For the Shared Task, the annotation scheme called "Universal Dependencies version 2", or "UD v2" for short will be used.

Participants will get UD treebanks in many languages, with raw text, gold-standard sentence and word segmentation, POS tags, dependency relations, and in many cases also lemmas + morphological features. The test data will contain none of the gold-standard annotations, but baseline predicted segmentation and POS tags will be available. Labeled attachment score (LAS) will be computed for every test set, and the macro-average of the scores over all test sets will provide the main system ranking.

The test sets will include a few **surprise languages.** We will not provide training data for these languages, only a small sample shortly before the test phase. To succeed in parsing these languages, systems will have to employ low-resource language techniques, utilizing data from other languages.

There will be no separate open and closed tracks. Instead, we will include every system in a single track, which will be formally closed, but where the list of permitted resources is rather broad and includes large raw corpora and parallel corpora.

Participating systems will have to find labeled syntactic dependencies between words, i.e. a syntactic *head* for each word, and a *label* classifying the type of the dependency relation. Participants will parse raw text where no gold-standard pre-processing (tokenization, lemmas, morphology) is available. However, there are at least two open-source pipelines (UDPipe, https://ufal.mff.cuni.cz/udpipe/, and SyntaxNet, https://www.tensorflow.org/versions/r0.9/tutorials/syntaxnet/index.html) that the participants can run instead of training their own models for any steps preceding the dependency analysis. We will even provide variants of the test data that have been preprocessed by UDPipe. We believe that this makes the task reasonably accessible.

The task is open to everyone. The organizers rely, as is usual in large shared tasks, on the honesty of all participants who might have some prior knowledge of part of the data that will eventually be used for evaluation, not to unfairly use such knowledge. The only exception is the chair of the organizing team, who cannot submit a system, and who will serve as an authority to resolve any disputes concerning ethical issues or completeness of system descriptions.

Timeline:

- December 11: Announcement of the shared task and set up of the shared task website. Registration for the Shared Task open.

- January 10: Deadline for suggesting additional data by participants (registration necessary)

- February 20 … trial data for a few languages will be available

- February 28: Task Registration deadline. Participants have to register to setup their evaluation space and other data, and get access to task data later.
- March 1 … training and development data will be released
- May 1 … surprise languages will be announced and small sample data released
- May 8 – 12 … test phase
- May 15 … results will be announced
- May 26 … submission of system description papers
- June 2 … reviews due
- June 9 … final papers due
- August 3 – 4 … CoNLL, Vancouver, Canada