ARL CrowdTree

Version 1.0

Contributors:

Stephen Tratz (ARL) and Nhien ‘Theresa’ Phan (CQL intern)

Purpose:

ARL CrowdTree is a browser-based graphical dependency tree editor. For a given sentence, users can modify the dependency tree structure as well as assign dependency labels (e.g., subject, direct object, indirect object) and part-of-speech labels (e.g., noun, verb). It is designed to be usable with Amazon’s Mechanical Turk crowdsourcing platform and is likely of greatest value to computational linguistics researchers who are running dependency tree annotation projects.

Large collections of dependency trees are useful for training and evaluating automatic dependency parsers, which don’t exist for the vast majority of the world’s languages and which are key components in many natural language processing workflows. By releasing this code open source, we aim to speed the development of such resources and to gain ARL recognition (and, perhaps, collaborators) within the greater computational linguistics research community.

Architecture:

The system consists of two components—a graphical frontend written entirely in HTML, CSS, and Javascript and a server-side Java servlet.

License:

The plan is to release the code under the Apache 2.0 license, which permits both commercial and non-commercial use, in order to maximize its potential impact.

Related publication:

Stephen Tratz and Nhien Phan. 2018. A Web-based System for Crowd-in-the-Loop Dependency Treebanking. In the Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). Miyazaki, Japan.

ARL CrowdTree screenshot