

CO504 Natural Language Processing

Autumn 2015

Assignment

Prepare a plan to implement Earley's parser for an Indian language. Identify all the intermediate steps required in the process and define the formats for input and output for each stage. Identify the resources required, such as, PoS tagset, lexicon and CFG, and define their formats and create initial versions of the resources. The linguistic resources and information should be in Unicode encoding. The plan should be language independent.

The exercise is to be carried out individually or in groups of two students. Deliverable should be a report (soft-copy and printout) and software (resources), and a presentation of 5-10 minutes.

[Submit by 26-10-2015].