

Introduction to NLP

Spring 2022, IIIT Hyderabad
Project Outline

Abhinav S Menon (2020114001)
Pratyaksh Gautam (2020114002)
Shashwat Singh (2020114016)

The Problem

We are developing a dependency parser for English. We will implement the system given here, which is deterministic apart from the oracle used. We have not yet decided how we will implement the oracle.

The system will take a sentence (in the form of a string) as input, and return its dependency graph. Thus tokenisation, POS tagging, and dependency parsing will be included in the pipeline.

The Method

The paper we have chosen describes a transition-based algorithm, which is deterministic except for the oracle.

The Scope

We will implement the system as outlined in the paper.

Timeline

A tentative timeline of the project is as follows:

- 26th February: Literature Review
- 5th March: Data Collection & Preprocessing
- 20th March: Training
- 31st March: Finalisation & Improvements