Engineering Standards and Requirements

Lightning Talk - SE 491
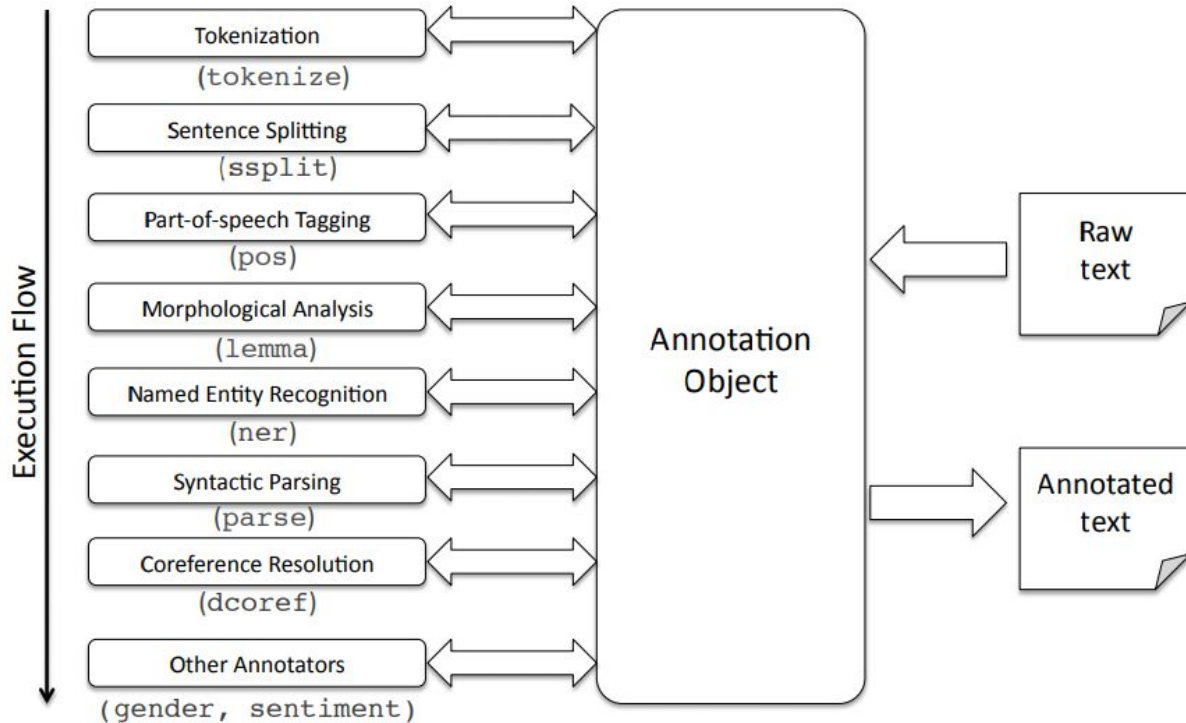**Team: sdmay21-35**
POS Tagger for Software Documentation

Ahmad Alramahi, Joseph Naberhaus

James Taylor & Ethan Ruchotzke, Austin Boling, Ekene Okeke

Stanford NLP

Java Vs. Python

- Figure out the optimal language for the Stanford NLP
- Java has high RAM usage making it undesirable currently
- We are working on lowering its RAM usage
- Python does not seem to have the same problems
- Python seems like a better candidate for our usage

Augmenting Stanford NLP

- To augment the Stanford NLP our group needs to do multiple things
- We need to learn how to efficiently compile and run the Stanford Tagging software
- We need to learn how to train the NLP to suit our needs
- We need to learn how to efficiently run the NLP, as it is demanding performance wise

Creating Training Data

- There are problems to work around
  - No existing dataset
  - Insufficient time to do all tagging manually
  - Tagging is relatively high skilled labor
- The solution will be to use the existing NLP model as a first-pass
- Then switch to manual tagging to correct errors
- Going to be a significant effort regardless

Project Roadmap