

**Base paper:** Identification of Rhetorical Roles of Sentences in Indian Legal Judgments

**Task:** Semantic Segmentation (sentence level)

### Why semantic segmentation of legal documents?

For sentence level labelling/tagging with semantic tags that can be used for further analysis with regulatory documents.

Current Approach: Semantic Segmentation
<ol style="list-style-type: none"><li>1. Define our own labels</li><li>2. Use the annotation approach &amp; methodology from the base paper ( BiLSTM CRF model)</li></ol>
Cons: <ol style="list-style-type: none"><li>1. Annotation of about 9k sentences from 50 relevant case files</li><li>2. Crawl 50k documents for sentence embedding pretraining</li></ol>
Next Step: Simplify extracted context/info and then map to regulation regulation

#### 1. Implementation:

##### a. Defined Labels

- i. Argument of accused
- ii. Findings by authority in charge
- iii. Context of violation
- iv. Regulatory text reference
- v. Issues raised
- vi. Others - not highlighted



Steps ahead - Improve labels, determine all sources of case files, annotate a sample doc from each source using current labels , incorporate feedback, finalize labels for all the data sources.

Q. Any alternative approach for semantics extraction from legal case files?

Q. Suggestions for improving labels?