Changes Made from Our Original Proposal

Dataset:

- Originally, we had planned to use DBpedia as the dataset for our project due to the consistent ontology. We envisioned employing rdf2hpcc to manipulate and interact with the RDF data files int the hpcc environment.
- The principal benefits associated with utilizing DBpedia data encompassed the abundance of existing datasets and DBpedia Live. DBpedia Live provides real time update to the DBpedia database so that our dataset is regularly updated.
- The main problem with this approach was DBpedia has very less data pertaining to the Indian Law.
- Only landmark supreme court decisions are available in the DBpedia database. Instead, we opted for using IDLC dataset. IDLC dataset contains over 35,000 supreme court judgements.
- Our dataset can be updated regularly by making use of API calls from https://indiankanoon.org website of which IDLC is a subset of. We also plan on scaling the project to include other datasets apart from IDLC.

Querying:

- While RDF offered a structured approach it is not inherently supported by hpcc, we planned on using Apache Jena and rdf2hpcc to convert it to xml and use xml version for querying.
- Whereas our current dataset is of csv and we can make use of existing hpcc functions for querying purposes.