# Updated Approach

Approach:

- The initial user interface will be a website where the user will enter a brief description of the case the person is fighting.
- In the website the keywords will be extracted with the help of an NLP model.(Note that the extracted keywords will be relevant to the context of law)
- These keywords will be sent to the HPCC cluster via a Roxie query.
- In the cluster we have our dataset sprayed, and a searching algorithm is used inorder to match all the extracted keywords and then to shortlist all the references with matching keywords.
- After this the shortlisted references will be ranked with the help of a ranking algorithm.
- Then this ranked set will be shown to the user.

Overview of the project:
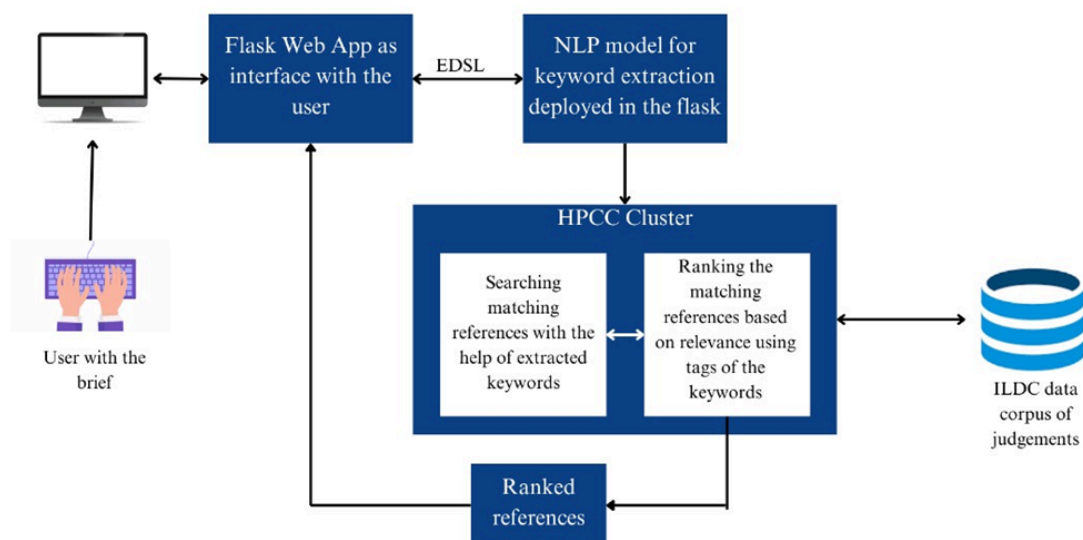
2. Difference from our previous approach

- In our previous approach we planned on using DBpedia as our data source, but now we decided our data source to IDLC supreme court judgments which has over 35,00 cases.
- Our current data source is of the format .csv and hence we won't be needing rdf2hpcc tools and also conversion of rdf to xml for querying purposes.

3. Justification

- Dbpedia has very few cases of Indian Law (less than 100), so to include more variety and also number of cases we changed our dataset to IDLC.

4. Contribution to the HPCC community:

1. Demonstration of interface between a Web-environment and HPCC cluster.
2. Development of a ranking algorithm to rank the shortlisted references.
3. Implementation of a searching algorithm.

New Timeline

Timeline	Task
December	Engage in the exploration of HPCC systems architecture and master the ECL programming language.
1st January - 1st February	Understand the working of NLP and create a web app interface for keyword extraction.
1st February - 13th February	Spray the ILDC dataset to Thor and develop a demo Roxie for searching data from predefined keywords.
13th February - 30th March	Roxie Query for searching data with endpoints to receive keywords from the web app.
30t March - 13th April	Integrate Web app and Roxie query using REST API.
13th April - 30th April	Develop a Ranking System.
30th April - 15th May	Integrate the web app with Roxie endpoint to receive ranked cases from the cluster.
15th May - 30th May	Refine and fine-tune the entire project through meticulous testing and optimization.
Till the end	Write Research paper.