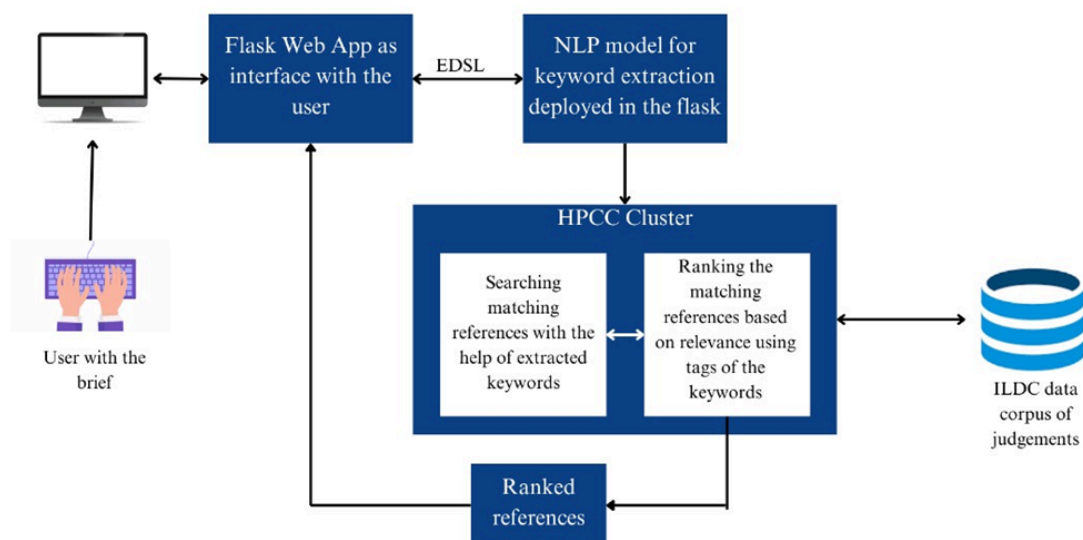


## Updated Approach

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.